

文章编号:1005-3085(2011)07-0001-42

## 2011 高教社杯全国大学生数学建模竞赛

谢金星

(清华大学, 北京 100084)

2011 年是全国大学生数学建模竞赛 20 周年。这项每年一次的竞赛由中国工业与应用数学学会于 1992 年创办, 从 1994 年起由教育部高教司和中国工业与应用数学学会共同主办, 2002 年起由高等教育出版社独家冠名赞助。在教育部领导“扩大受益面, 保证公正性, 推动教育改革”的指示精神指导下, 在各级教育行政部门和广大教师的积极指导和参与下, 二十年来这项竞赛发展迅速。20 年前开始举办这项竞赛时, 只有 74 所高校的 314 队参加; 到今年, 参赛学校数达到 1251 所, 超过了我国现有高校总数的一半; 参赛队数达 19490 队 (参赛学生 58000 多人), 参赛学校数和队数分别每年平均以超过 16% 及 24% 的速度增长。数学建模竞赛不仅已成为我国高校历史最悠久、规模最大、影响广泛的学科性竞赛活动, 而且成功地推动了数学教学的改革, 造就了这些年来规模最大也最成功的、以推进数学建模为核心的数学教学改革实践。

2011 高教社杯全国大学生数学建模竞赛 (即第 20 届竞赛) 于 2011 年 9 月 9 日至 12 日举行, 来自全国 33 个省/市/自治区 (包括香港和澳门特区) 以及新加坡和美国的 1251 所高校 19490 个队 (其中本科组 16008 队、专科组 3248 队) 的近六万多名大学生参加了今年的比赛, 是历年来参赛人数最多的, 与 2010 年的 1196 所院校的 17317 队 (其中本科组 14119 队、专科组 3198 队) 相比, 校数增长 4.6%, 队数增长 12.5% (本专科队数分别增长 13.4% 和 8.9%)。竞赛继续得到了北京迈斯沃克软件有限公司的赞助与支持, 并设立了 Matlab 创新奖。

竞赛答卷首先在由省/市/自治区形成的 27 个赛区和联合赛区 (今年由内蒙、宁夏、青海、西藏组成) 进行初评, 评出各赛区的获奖者, 然后各赛区按一定比例将优秀答卷送全国竞赛组委会 (与去年一样, 对境外参赛者来说, 本项竞赛也被称为“当代大学生数学建模竞赛”, 论文单独进行评奖)。送全国评阅的论文总数是 1821 份 (A、B、C、D 题分别为 1037, 463, 204, 117 份), 约占报名参赛队数的 9%。A 题还有 1 篇按创新点论文送评。全国组委会聘请了 48 位专家组成全国评阅专家组, 认真评阅了候选论文, 也对一些队进行了面试答辩。经过全国评阅专家的评阅与全国组委会的复核、面试, 以及异议期的审核, 按照从严控制获奖比例特别是全国一等奖数量的原则, 确定获全国奖共 1536 队, 获奖比例约为 8%。与去年相比基本持平, 其中本科组一等奖 224 队, 二等奖 1040 队, 专科组一等奖 51 队, 二等奖 221 队。一、二等奖分别占参赛总数的 1.5% 和 6.5%。高教社杯的获得者本科组为南京信息工程大学黄思、盛振峰、周敏强, 专科组为江西应用技术职业学院杨忠、张岐良、徐小辉。Matlab 创新奖获得者本科组为第三军医大学唐棣、董小小、魏歆, 专科组为九江学院乔衡山、许琼、徐景华。同时, 北京、天津、山西、上海、江苏、浙江、山东、广东、重庆、四川十个赛区组委会获得优秀组织工作奖。

这项竞赛之所以受到大学生们如此热烈的欢迎, 是因为它有以下的特点: 赛题由工程技术、管理科学、社会生活等领域的实际问题简化加工而成, 要求参赛者结合实际问题灵活运用数学和计算机软件以及其他学科的知识, 通过建立、求解、评估、改善数学模型, 充分发挥其

聪明才智和创造精神;三名大学生组成一队,团结合作,选择一题在三天时间内完成一篇研究论文;可以自由地收集、查阅资料,调查研究,使用计算机、互联网和各种软件(但是不能与队外的任何人讨论赛题);赛题没有事先确定的答案,论文评阅的标准是:假设的合理性、建模的创造性、结果的正确性和表述的清晰性。

为了进一步加强竞赛规则的执行,促进竞赛的健康发展,要求参赛同学签名承诺:“我们完全明白,在竞赛开始后参赛队员不能以任何方式(包括电话、电子邮件、网上咨询等)与队外的任何人(包括指导教师)研究、讨论与赛题有关的问题。我们郑重承诺,严格遵守竞赛纪律,以保证竞赛的公平、公正性。如有违反竞赛规则的行为,我们将会受到严肃处理”。

今年A题“城市表层土壤重金属污染分析”由山东理工大学李功胜教授提出,B题“交巡警服务平台的设置与调度”由解放军后勤工程学院但琦教授和解放军信息工程大学韩中庚教授提出,C题“企业退休职工养老金制度的改革”由济南大学许振宇教授提出,D题“天然肠衣搭配问题”由复旦大学陆立强教授提出,在此一并表示感谢。为了更广泛、有效地收集适合竞赛的题目和素材,全国大学生数学建模竞赛组委会再次向全社会诚征赛题,有意者请与全国组委会副秘书长蔡志杰联系,将赛题或素材用电子邮件直接发给蔡志杰教授(E-mail:zhijiecai@163.com),并请在邮件主题(Subject)中注明“赛题素材”,在邮件正文中写明单位、姓名、电话等信息,以便联系。同时,务必注意保密,切勿将与素材有关的任何内容与其他人交流。

为了与广大师生进行交流,对今后的竞赛予以适当引导,全国评阅专家组选择了一些优秀论文在本刊发表(包括新加坡国立大学的一篇参加“当代大学生数学建模竞赛”的获奖论文),并请命题者或评阅者撰文讲评。

发表的论文是同学们三天内写出的,为了保持原貌只作了适当的删节和文字上的修正,文章不可避免地存在着相当多的不妥之处,请读者谅解。

希望对这项竞赛作进一步了解的读者,请访问竞赛网站:<http://mcm.edu.cn>。

## A题 城市表层土壤重金属污染分析

随着城市经济的快速发展和城市人口的不断增加,人类活动对城市环境质量的影响日显突出,对城市土地地质环境异常的查证,以及如何应用查证获得的海量数据资料开展城市环境质量评价,研究人类活动影响下城市地质环境的演变模式,日益成为人们关注的焦点。

按照功能划分,城区一般可分为生活区、工业区、山区、主干道路区及公园绿地区等,分别记为1类区、2类区、...、5类区,不同的区域环境受人类活动影响的程度不同。

现对某城市城区土地地质环境进行调查,为此,将所考察的城区划分为间距1公里左右的网格子区域,按照每平方公里1个采样点对表层土(0~10厘米深度)进行取样、编号,并用GPS记录采样点的位置。应用专门仪器测试分析,获得了每个样本所含的多种化学元素的浓度数据。另一方面,按照2公里的间距在那些远离人群及工业活动的自然区取样,将其作为该城区表层土壤中元素的背景值。

附件1列出了采样点的位置、海拔高度及其所属功能区等信息,附件2列出了8种主要重金属元素在采样点处的浓度,附件3列出了8种主要重金属元素的背景值。

现要求你们通过数学建模来完成以下任务:

1) 给出8种主要重金属元素在该城区的空间分布,并分析该城区内不同区域重金属的污染程度;

2) 通过数据分析,说明重金属污染的主要原因;

3) 分析重金属污染物的传播特征, 由此建立模型, 确定污染源的位置;

4) 分析你所建立模型的优缺点, 为更好地研究城市地质环境的演变模式, 还应收集什么信息? 有了这些信息, 如何建立模型解决问题?

注: 附件数据略, 请从 <http://mcm.edu.cn> 下载.

## B题 交巡警服务平台的设置与调度

“有困难找警察”, 是家喻户晓的一句流行语. 警察肩负着刑事执法、治安管理、交通管理、服务群众四大职能. 为了更有效地贯彻实施这些职能, 需要在市区的一些交通要道和重要部位设置交巡警服务平台. 每个交巡警服务平台的职能和警力配备基本相同. 由于警务资源是有限的, 如何根据城市的实际情况与需求合理地设置交巡警服务平台、分配各平台的管辖范围、调度警务资源是警务部门面临的一个实际课题.

试就某市设置交巡警服务平台的相关情况, 建立数学模型分析研究下面的问题:

1) 附件1中的附图1给出了该市中心城区A的交通网络和现有的20个交巡警服务平台的设置情况示意图. 相关的数据信息见附件2. 请为各交巡警服务平台分配管辖范围, 使其在所管辖的范围内出现突发事件时, 尽量能在3分钟内有交巡警(警车的时速为60 km/h)到达事发地点.

对于重大突发事件, 需要调度全区20个交巡警服务平台的警力资源, 对进出该区的13条交通要道实现快速全封锁. 实际中一个平台的警力最多封锁一个路口, 请给出该区交巡警服务平台警力合理的调度方案.

根据现有交巡警服务平台的工作量不均衡和有些地方出警时间过长的实际情况, 拟在该区内再增加2至5个平台, 请确定需要增加平台的具体个数和位置.

2) 针对全市(主城六区A, B, C, D, E, F)的具体情况, 按照设置交巡警服务平台的原则和任务, 分析研究该市现有交巡警服务平台设置方案(参见附件)的合理性. 如果有明显不合理, 请给出解决方案.

如果该市地点P(第32个节点)处发生了重大刑事案件, 在案发3分钟后接到报警, 犯罪嫌疑人已驾车逃跑. 为了快速搜捕嫌疑犯, 请给出调度全市交巡警服务平台警力资源的最佳围堵方案.

附件1: A区和全市六区交通网络与平台设置的示意图.

附件2: 全市六区交通网络与平台设置的相关数据表(共5个工作表).

注: 附件数据略, 请从 <http://mcm.edu.cn> 下载.

## C题 企业退休职工养老金制度的改革

养老金也称退休金, 是一种根据劳动者对社会所作贡献及其所具备享受养老保险的资格, 以货币形式支付的保险待遇, 用于保障职工退休后的基本生活需要.

我国企业职工基本养老保险实行“社会统筹”与“个人账户”相结合的模式. 即企业把职工工资总额按一定比例(20%)缴纳到社会统筹基金账户, 再把职工个人工资按一定比例(8%)缴纳到个人账户. 这两个账户我们合称为养老保险基金. 退休后, 按职工在职期间每月(或年)的缴费工资与社会平均工资之比(缴费指数), 再考虑到退休前一年的社会平均工资等因素, 从社会统筹账户中拨出资金(基础养老金), 加上个人工资账户中一定比例的资金(个人账户养老金), 作为退休后每个月的养老金. 养老金会随着社会平均工资的调整而调整. 如果

职工死亡, 社会统筹账户中的资金不退给职工, 个人账户中的余额可继承, 个人账户储存额以银行当时公布的一年期存款利率计息, 为简单起见, 利率统一设定为3%。

养老金的发放与职工在职时的工资及社会平均工资有着密切关系; 工资的增长又与经济增长相关。近30年来我国经济发展迅速, 工资增长率也较高; 而发达国家的经济和工资增长率都较低。我国经济发展的战略目标, 是要在21世纪中叶使我国人均国民生产总值达到中等发达国家水平。

现在我国养老保险改革正处于过渡期, 养老保险管理的一个重要的目标是养老保险基金的收支平衡, 它关系到社会稳定和老龄化社会的顺利过渡。影响养老保险基金收支平衡的一个重要因素是替代率, 替代率是指职工刚退休时的养老金占退休前工资的比例。按照国家对基本养老保险制度的总体思路, 未来基本养老保险的目标替代率确定为58.5%。替代率较低, 退休职工的生活水准低, 养老保险基金收支平衡容易维持; 替代率较高, 退休职工的生活水准就高, 养老保险基金收支平衡较难维持, 可能出现缺口。所谓缺口, 是指当养老保险基金入不敷出时出现的收支之差。

附件1是山东省职工历年平均工资数据; 附件2是2009年山东省某企业各年龄段职工的工资分布情况, 附件3是养老金的计算办法。请建立数学模型, 解决如下问题:

问题一: 对未来中国经济发展和工资增长的形势做出你认为是简化、合理的假设, 并参考附件1, 预测从2011年至2035年的山东省职工的年平均工资。

问题二: 根据附件2计算2009年该企业各年龄段职工工资与该企业平均工资之比, 如果把此比值看作职工缴费指数的参考值, 考虑该企业职工自2000年起分别从30岁、40岁开始缴养老保险, 一直缴费到退休(55岁, 60岁, 65岁), 计算各种情况下的养老金替代率。

问题三: 假设该企业某职工自2000年起从30岁开始缴养老保险, 一直缴费到退休(55岁, 60岁, 65岁), 并从退休后一直领取养老金至75岁死亡。计算养老保险基金的缺口情况, 并计算该职工领取养老金到多少岁时, 其缴存的养老保险基金与其领取的养老金之间达到收支平衡。

问题四: 如果既要达到目标替代率, 又要维持养老保险基金收支平衡, 你认为可以采取什么措施。请给出你的理由。

注: 附件数据略, 请从<http://mcm.edu.cn>下载。

## D 题 天然肠衣搭配问题

天然肠衣(以下简称肠衣)制作加工是我国的一个传统产业, 出口量占世界首位。肠衣经过清洗整理后被分割成长度不等的小段(原料), 进入组装工序。传统的生产方式依靠人工, 边丈量原料长度边心算, 将原材料按指定根数和总长度组装出成品(捆)。

原料按长度分档, 通常以0.5米为一档, 如: 3米-3.4米按3米计算, 3.5米-3.9米按3.5米计算, 其余的依此类推。表1是几种常见成品的规格, 长度单位为米,  $\infty$ 表示没有上限, 但实际长度小于26米。

为了提高生产效率, 公司计划改变组装工艺, 先丈量所有原料建立一个原料表, 表2为某批次原料描述。根据成品和原料描述, 设计一个原料搭配方案, 工人根据这个方案“照方抓药”进行生产。

公司对搭配方案有以下具体要求:

- 1) 对于给定的一批原料, 装出的成品捆数越多越好;
- 2) 对于成品捆数相同的方案, 最短长度最长的成品越多, 方案越好;
- 3) 为提高原料使用率, 总长度允许有 $\pm 0.5$ 米的误差, 总根数允许比标准少1根;

4) 某种规格对应原料如果出现剩余, 可以降级使用。如长度为 14 米的原料可以和长度介于 7 米-13.5 米的进行捆扎, 成品属于 7 米-13.5 米的规格;

5) 为了食品保鲜, 要求在 30 分钟内产生方案。

请建立上述问题的数学模型, 给出求解方法, 并对表 1 和表 2 给出的实际数据进行求解, 给出搭配方案。

表 1: 成品规格表

最短长度(米)	最大长度(米)	根数	总长度(米)
3	6.5	20	89
7	13.5	8	89
14	$\infty$	5	89

表 2: 原料描述表

长度(米)	3-3.4	3.5-3.9	4-4.4	4.5-4.9	5-5.4	5.5-5.9	6-6.4	6.5-6.9
根数	43	59	39	41	27	28	34	21
长度(米)	7-7.4	7.5-7.9	8-8.4	8.5-8.9	9-9.4	9.5-9.9	10-10.4	10.5-10.9
根数	24	24	20	25	21	23	21	18
长度(米)	11-11.4	11.5-11.9	12-12.4	12.5-12.9	13-13.4	13.5-13.9	14-14.4	14.5-14.9
根数	31	23	22	59	18	25	35	29
长度(米)	15-15.4	15.5-15.9	16-16.4	16.5-16.9	17-17.4	17.5-19.9	18-18.4	18.5-18.9
根数	30	42	28	42	45	49	50	64
长度(米)	19-19.4	19.5-19.9	20-20.4	20.5-20.9	21-21.4	21.5-21.9	22-22.4	22.5-22.9
根数	52	63	49	35	27	16	12	2
长度(米)	23-23.4	23.5-23.9	24-24.4	24.5-24.9	25-25.4	25.5-25.9		
根数	0	6	0	0	0	1		

## 2011 高教社杯全国大学生数学建模竞赛获奖名单

高教社杯获得者: 本科组: 黄 思、盛振峰、周敏强(南京信息工程大学)  
专科组: 杨 忠、张岐良、徐小辉(江西应用技术职业学院)

Matlab 创新奖获得者: 本科组: 唐 楝、董小小、魏 歆(第三军医大学)  
专科组: 乔衡山、许 琼、徐景华(九江学院)



文章编号:1005-3085(2011)07-0043-07

## 表层土壤中重金属污染分析

蔡志杰<sup>1,2</sup>

(1- 复旦大学数学科学学院, 上海 200433; 2- 上海市现代应用数学重点实验室, 上海 200433)

摘 要: 本文对 2011 年全国大学生数学建模竞赛 A 题进行分析, 给出各问题的求解方法, 并对学生的论文进行评述.

关键词: 重金属污染; 散乱数据插值; 评价指标; 对流方程; 特征线法

分类号: AMS(2000) 65D05; 62H12; 80A23; 62P12 中图分类号: O241.3; O212.4; X8 文献标识码: A

### 1 问题的提出

2011 年全国大学生数学建模竞赛 A 题“表层土壤中重金属污染分析”来自于环境监测部门的一个实际问题, 题目给出了一组采样点上 8 种重金属元素的浓度值, 要求作出污染程度的评价、分析污染原因、确定污染源的位置, 本文给出各问题的求解方法, 并对评阅中发现的问题进行分析与评述.

### 2 地形高度函数和浓度分布函数的插值方法

首先观察各采样点的分布情况, 如图 1 所示, 其中“●”为生活区, “◆”为工业区, “■”为山区, “▼”为交通区, “★”为公园绿地区. 各采样点不是分布在规则的网格点上, 而是呈现出散乱状态, 因而应采用散乱数据插值.

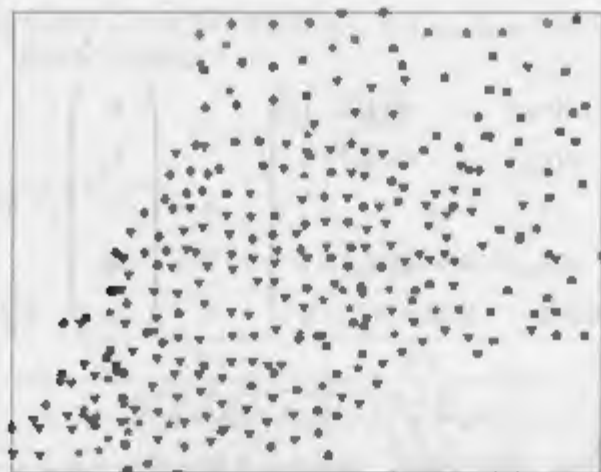


图 1: 采样点分布图

一种插值方法是 Shepard 插值<sup>[1,2]</sup>, 其基本思想是: 某一点的函数值受周围各点的影响, 较近的点影响较大, 较远的点影响较小, 其影响权数与距离平方成反比, 因而又称为距离平方反比律, 具体方法如下:

设采样点为  $P_i(x_i, y_i)$ , 其高度为  $h_i$  ( $i = 1, 2, \dots, n$ ), 其中  $n$  为采样点数, 设地形高度函数为  $h(x, y)$ , 则对任一点  $P(x, y)$ , 其高度为

$$h(x, y) = \begin{cases} h_j, & \text{当某个 } r_j = 0 \text{ 时,} \\ \sum_{i=1}^n \frac{h_i}{r_i^2} / \sum_{i=1}^n \frac{1}{r_i^2}, & \text{其它,} \end{cases} \quad (1)$$

其中

$$r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (2)$$

为  $P$  点与  $P_i$  点之间的距离,

另一种常用的散乱数据插值方法是 Kriging 插值<sup>[2]</sup>, 它建立在地统计学理论上, 利用区域化变量的原始数据和半方差函数的结构特征, 对位采样点的区域化变量的取值进行线性最佳无偏估计,

设任一点  $P(x, y)$  的高度为

$$h(x, y) = m + \sum_{i=1}^n \lambda_i \phi(r_i), \quad (3)$$

其中  $P$  点与  $P_i$  点之间的距离  $r_i$  由 (2) 式给出,  $m$  和  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) 为待定系数, 满足

$$\sum_{i=1}^n \lambda_i = 1 \quad (4)$$

及插值条件

$$h(x_i, y_i) = h_i, \quad i = 1, 2, \dots, n. \quad (5)$$

由此得到关于  $m$  和  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) 的线性代数方程组

$$AA^T = b, \quad (6)$$

其中

$$A = \begin{pmatrix} \phi(r_{11}) & \phi(r_{12}) & \cdots & \phi(r_{1n}) & 1 \\ \phi(r_{21}) & \phi(r_{22}) & \cdots & \phi(r_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi(r_{n1}) & \phi(r_{n2}) & \cdots & \phi(r_{nn}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{pmatrix}, \quad b = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \\ 1 \end{pmatrix}, \quad (7)$$

这里

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (8)$$

为  $P_i$  点与  $P_j$  点之间的距离, 解此方程组, 得到待定系数  $m$  和  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) 的值, 然后利用插值函数 (3) 即可得到高度函数  $h(x, y)$ .

用同样的方法对重金属元素的浓度进行插值, 可以得到各重金属元素的浓度分布函数.

3 污染程度评价指标

土壤重金属污染评价的方法很多，如地质累积指数<sup>[3,4]</sup>、内梅罗指数法<sup>[5]</sup>等。

3.1 地质累积指数

地质累积指数，又称为 Muller 指数，是 Muller 于 1949 年提出的针对沉积物种的重金属污染程度的评价方法。它不仅考虑了自然地质过程造成的背景值影响，而且也充分注意了人为活动对重金属污染的影响，因此，这一指数不仅反映了重金属分布的自然变化特征，而且可以判别人为活动对环境的影响，是区分人为活动影响的重要参数<sup>[3,4]</sup>。地质累积指数定义为

$$I_{geo} = \log_2 \frac{c}{k\mu}, \tag{9}$$

其中  $c$  为重金属元素浓度的实际测量值， $\mu$  为重金属元素的背景值， $k$  为修正指数，通常取为 1.5。

不同的地质累积指数  $I_{geo}$  对应于不同的污染程度，其分级标准参见表 1<sup>[3]</sup>。

表 1：地质累积指数分级表

地质累积指数	地质累积指数分级	污染程度
$I_{geo} < 0$	0	无污染
$0 \leq I_{geo} < 1$	1	无污染-中度污染
$1 \leq I_{geo} < 2$	2	中度污染
$2 \leq I_{geo} < 3$	3	中度污染-强污染
$3 \leq I_{geo} < 4$	4	强污染
$4 \leq I_{geo} < 5$	5	强污染-极强污染
$I_{geo} \geq 5$	6	极强污染

3.2 内梅罗指数

为了评价各功能区重金属污染的程度，需要将各单个指标进行合成，即构造综合指标。常用的环境质量综合评价指标是内梅罗指数<sup>[5]</sup>。这是一种兼顾极值或称突出最大值的加权型多因子环境质量指数，即特别考虑了污染最严重的因子。记  $I_{ave}$ 、 $I_{max}$  分别为各单因子环境质量指数的平均值和最大值，则内梅罗指数定义为

$$P = \sqrt{\frac{I_{ave}^2 + I_{max}^2}{2}}. \tag{10}$$

污染指数分级方法如表 2 所示<sup>[6]</sup>。

表 2：内梅罗综合污染指数的分级标准

污染等级	$P$	污染描述	污染程度
I	$P < 0.7$	安全	清洁
II	$0.7 \leq P < 1.0$	警戒线	尚清洁
III	$1.0 \leq P < 2.0$	轻污染	超标
IV	$2.0 \leq P < 3.0$	中污染	中度污染
V	$P \geq 3.0$	重污染	严重污染



## 4 污染原因分析

### 4.1 相关性分析

为了分析污染原因,考虑不同重金属元素之间是否存在一定的联系.如果不同的重金属存在很高的相关性,那么它们可能存在共生性,因而有理由相信,它们是由同一污染源产生的,从而可以分析出污染产生的原因.

相关系数定义为

$$\rho = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}. \quad (11)$$

### 4.2 聚类分析

相关系数仅考虑重金属元素两两之间的关系,要研究多种重金属元素之间的关系,可采用聚类分析<sup>[7]</sup>的方法.

聚类分析方法是基于数值分类法的思想建立起来的,又称为系统聚类法,它是研究样品或指标分类的一种多元统计分析方法.这里是基于样本指标进行聚类,称为R型聚类,其步骤为:

步骤1 将各重金属元素视为不同的 $n$ 类,每一类包含一个元素;

步骤2 计算各指标两两之间的相关性 $\rho(x_i, x_j)$  ( $i, j = 1, \dots, n$ );

步骤3 将相关性最大的两类合并成一类;

步骤4 计算新合成的类与其他各类之间的相关系数;

步骤5 重复步骤3和步骤4,直到将元素合并成我们所需要的类.

两类之间的相关系数可以用广义相关系数<sup>[8]</sup>来计算.设 $a_1, a_2, \dots, a_m$ 和 $b_1, b_2, \dots, b_n$ 为给定的两组 $p$ 维列向量,记

$$x = (a_1, a_2, \dots, a_m)_{p \times m}, \quad y = (b_1, b_2, \dots, b_n)_{p \times n}. \quad (12)$$

令

$$M_{xx} = E(x - Ex)^T(x - Ex), \quad M_{yy} = E(y - Ey)^T(y - Ey), \quad (13)$$

$$M_{yx}^T = M_{xy} = E(x - Ex)^T(y - Ey). \quad (14)$$

用 $\lambda_1, \lambda_2, \dots, \lambda_l$ 表示 $M_{xx}^+ M_{xy} M_{yy}^+ M_{yx}$ 的全部非零特征根,其中 $A^+$ 表示矩阵 $A$ 的广义逆.定义 $x$ 与 $y$ 之间的广义相关系数为

$$\rho^*(x, y) = \frac{1}{l} \sum_{i=1}^l \lambda_i. \quad (15)$$

可以证明:当 $m = n = 1$ 时,广义相关系数即为相关系数的平方<sup>[8]</sup>,即 $\rho^* = \rho^2$ .

## 5 污染源位置的确定

为了确定污染源位置,首先需要分析污染物传播的特征.由于我们主要考虑表层土壤,地表径流冲刷为主要传播途径,因而对流是最主要传播特征,而扩散所起的作用相对来说较小.

记 $u = u(x, y, t)$ 为 $t$ 时刻位于 $P(x, y)$ 点处重金属元素的浓度,利用质量守恒定律,可以得到对流方程<sup>[9]</sup>

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{q} = 0. \quad (16)$$

由于重金属元素的传播主要是由于雨水冲刷引起的, 而雨水只能从高处往低处流动, 因而可假设流量  $q$  与地形海拔高度  $h$  的梯度成比例, 即  $q \propto \nabla h$ .

此外, 假设流量  $q$  与污染元素浓度成正比应是合理的, 即  $q \propto u$ , 因此, 可令

$$q = -ku\nabla h, \quad (17)$$

这里负号表示污染元素从高处随雨水向低处传播,  $k (> 0)$  为比例系数.

将 (17) 式代入 (16) 式, 得到

$$\frac{\partial u}{\partial t} - k\nabla h \cdot \nabla u - ku\nabla^2 h = 0. \quad (18)$$

假设时刻  $T$  测得的各观测点处污染元素的浓度为  $u_i$  ( $i = 1, 2, \dots, n$ ), 则初始条件可写为

$$t = T: u(x_i, y_i) = u_i, \quad i = 1, 2, \dots, n. \quad (19)$$

由此得到了重金属污染的传播模型 (18)–(19), 模型的具体推导过程可参见文献 [10], 这是一个一阶双曲型方程, 可以通过特征线法求解.

由于只有一个时间点上的截面数据, 无法确定系数  $k$  的值, 因此对时间  $t$  作伸缩变换, 即令  $t' = kt$ , 则方程 (18) 可改写为 (为方便起见, 仍记  $t'$  为  $t$ )

$$\frac{\partial u}{\partial t} - \nabla h \cdot \nabla u - u\nabla^2 h = 0. \quad (20)$$

与 (18) 相比, 实际上是令  $k = 1$ .

公式 (20) 的特征线满足

$$\frac{dx}{dt} = -\nabla h, \quad (21)$$

若给定初始条件

$$t = T: x = x_T, \quad (22)$$

则记过给定点  $(T, x_T)$  的特征线为  $x(t; x_T)$ . 沿着特征线, 重金属元素浓度  $u$  满足

$$\frac{du}{dt} = u\nabla^2 h, \quad (23)$$

且初始条件为

$$t = T: u = u_T, \quad (24)$$

从而沿特征线的解为

$$u(x, t) = u_T \exp \left( - \int_t^T \nabla^2 h(x(\tau; x_T)) d\tau \right). \quad (25)$$

对 (21) 和 (25) 分别作离散化, 得到迭代格式 (注意, 这里沿时间  $t$  反向计算)

$$x^{m+1} = x^m + \nabla h(x^m) \Delta t, \quad (26)$$

$$u^{m+1} = u^0 \exp \left( - \sum_{j=1}^m \nabla^2 h(x^j) \Delta t \right), \quad (27)$$

其中  $u^0$  为  $T$  时刻污染元素的初始浓度.

若对某一定点  $x^0$  的浓度为  $u^0$ , 则由(26)及(27)可以求出各时刻  $t(t < T)$  沿着经过点  $x^0$  的特征线上的浓度值  $u(x(t), x^0)$ . 这样, 当给定污染浓度的阈值  $U$  时, 对给定时刻  $t(t \leq T)$  浓度超过阈值  $U$  的区域为

$$B_t = \{x(t) | u(x(t), t; x^0) \geq U\}, \quad (28)$$

此区域即为可能的污染源所在的区域.

## 6 评阅中发现的问题

在今年评阅过程中, 我们发现同学们大多能利用网络搜索资料, 查阅了有关的参考文献, 对数学软件(如 Matlab 等)使用较为熟练. 但同时也发现了一些问题.

1) 数据的插值拟合. 绝大多数参赛队都采用 MATLAB 软件来处理, 但大多数队没有说明应该用什么方法, 为什么用这种方法. 多数队用了 Matlab 中的 `griddata` 函数来做拟合, 该函数提供了 4 种插值方法, 而缺省方法为分块线性插值. 分块线性插值得到的结果最主要的问题是不光滑, 在后续处理时, 无法求导, 因此这是一种不太好的方法.

目前数学软件的功能非常强大, 可以很方便地利用数学软件来处理我们的问题. 但应注意, 在使用时应该先了解所使用的方法, 特别注意该方法的适用范围, 能否解决我们的问题, 有什么局限性, 可能会呈现什么问题, 然后有针对性地进行应用.

2) 对评价指标的选择. 有些队只给出了单个指标, 而未作综合处理. 有些队仅对单个指标采用简单的比例处理. 这样, 对污染评价及对整个区域的综合评价缺乏有力的数据支持, 得到的结论比较粗糙.

绝大多数队都从网上查找现成的指标, 但很少有同学深入研究这些指标. 例如地质量指数. 这一指数未能充分利用问题中给出的信息, 是题目中除了给出各重金属元素背景值的平均值, 同时还给出了其标准差及范围. 简单计算可知, 其结果的上下限分别为  $\mu_j \pm 2\sigma_j$ , 其中  $\mu_j$  为第  $j$  种重金属元素背景值的平均值,  $\sigma_j$  为相应的标准差. 因此在构造评价指标时, 应该将标准差考虑进去.

一种改进的方法为对数据作标准化处理

$$\tilde{c} = \frac{c - \mu}{\sigma}, \quad (29)$$

其中  $c$  为污染浓度的测量值,  $\mu, \sigma$  分别为背景值的平均值与标准差, 则改进的污染指数为

$$I = \log_2 \max\{\tilde{c} - 1, 1\} \quad (30)$$

3) 相关性分析也可以用秩相关系数, 方法是对两个指标分别按大小进行排序, 记其序号分别为  $1, 2, \dots, n$ , 则秩相关系数定义为

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum d_i^2 \quad (31)$$

其中  $d_i$  为指标之间的序号之差.

4) 污染元素的传播还可以用对流扩散模型与扩散模型建模. 虽然我们已经说明了, 在这个问题中扩散效应是很小的, 但在确定污染源时, 很多队假设污染浓度服从高斯分布, 这实际上是用了一个污染源的扩散模型, 这一模型的解就是高斯分布. 但对实测数据作正态性检验发现, 污染浓度并不是正态的. 因此, 应该考虑多个污染源的情况.

有的队在确定污染源时,对数据作了仔细的分析,发现污染浓度随着地形高度的增加是递减的,经回归分析发现,是指数衰减的。将高度影响剔除后,再用高斯-态分布的迭加形式进行搜索,这样得到的结果比较有说服力。

有的队在数据处理之前去除了浓度的异常值,这是不对的。异常值附近极有可能是污染源所在之处,将其去除,整体变得平缓了,污染源的位置也就变得不清晰了。

5) 当然,由于信息不足,题目只给出了一个时间上的截面数据,没有时间方向上的序列数据。因而要精确确定污染源的位置是不可能的,需要作进一步的数据采集工作。但这是实际工作中经常碰到的情况。采集数据需要大量的人力、物力、财力,而且第一次采样前的数据是无法获得的。数学模型的重要作用就是在有限数据,甚至数据不足、有误差的情况下,给出合理的结果,以指导今后的工作。当然,数学模型也有其局限性,只能起到辅助的作用。

## 参考文献

1. 田玉琴. 数学建模竞赛辅导教材[M]. 南京: 南京教育出版, 1997.  
Ye Y Q. Teaching Materials for Undergraduate Mathematical Contest in Modeling (2, M). Hunan: Hunan Education Press, 1997.
2. 王元敏. 散乱数据拟合的模型[M]. 数学理论[M]. 北京: 科学出版社, 2007.  
Wu Z M. Minimization Methods and Theories of Scattered Data Fitting[M]. Beijing: Science Press, 2007.
3. 百度科技. baidu baidu.com/view/5817487.htm, 2011.  
Baidu Encyclopedia. <http://baidu.baidu.com/view/5817487.htm>, 2011.
4. 蔡志杰. 地形指数法在土壤重金属污染评价中的应用[J]. 同济大学学报(自然科学版), 2006, 34(12): 1657-1660.  
Cai Z J. W. Applications of cumulative index method in the evaluation of non-heavy metal pollution[J]. Journal of Tongji University (Natural Science), 2006, 34(12): 1657-1661.
5. 百度百科. baidu baidu.com/view/4883176.htm, 2011.  
Baidu Encyclopedia. <http://baidu.baidu.com/view/4883176.htm>, 2011.
6. 中国大学网. <http://wenku.baidu.com/view/a2f5e04ac45e3b3567ec8b9f.html>, 2011.  
China University. <http://wenku.baidu.com/view/a2f5e04ac45e3b3567ec8b9f.html>, 2011.
7. Anderson T W. An Introduction to Multivariate Statistical Analysis[M]. New York: Wiley-Interscience, 2003.
8. 张北经. 文相关系数及其应用[J]. 数学学报, 1978, 1(4): 312-320.  
Zhang Y T. Generalized correlation coefficient and its applications[J]. Acta Mathematica Applicata Sinica, 1978, 1(4): 312-320.
9. 陈维雄. 初等微分方程[M]. 北京: 科学出版社, 2008.  
Chen W X. et al. Applied Partial Differential Equation[M]. Beijing: Science Press, 2008.
10. 蔡志杰, 龙敏. 城市表层土壤中重金属污染的传播模型[J]. 数学建模及其应用, 2012, 1(1): 53-56.  
Cai Z J. Long M B. Propagation model of heavy metals contamination in urban topsoil[J]. Mathematical Modeling and its Applications, 2012, 1(1): 53-56.

## Analysis of Heavy Metals Contamination in Urban Topsoil

CAI Zhi-jie<sup>1,2</sup>

(1- School of Mathematical Sciences, Fudan University, Shanghai, 200433)

(2- Shanghai Key Laboratory of Modern Applied Mathematics, Shanghai, 200433)

**Abstract:** In this paper, we discuss Problem A in CCMCM 2011 and present the methods of solving these problems. We also give some comments on the students' papers.

**Keywords:** heavy metal contamination, interpolation for scattering data, evaluation index, convection equation, character method

文章编号 1005-3085(2011)07-0050-09

## 城区重金属污染评价及源识别模型研究

齐 谦 成生群, 燕斌斌

指导教师: 刘保东

(山东大学, 济南 250100)

**摘 要** 本文主要研究基于多元统计分析的污染源识别模型。首先, 对多元统计分析中的主成分分析、聚类分析、判别分析等方法进行了简要介绍。然后, 基于多元统计分析方法, 建立了污染源识别模型。最后, 通过实例验证了该模型的有效性。结果表明, 该模型能够有效地识别污染源, 为环境管理部门提供决策依据。

**关 键 词** 多元统计分析; 污染源识别; 判别分析; 聚类分析; 主成分分析

**分 类 号** AMS: 2010 62A15 62H10 62P10 中图分类号: X28 122 文献标识码: A

### 1 问题提出

随着工业化和城市化的快速发展, 环境问题日益突出。重金属污染作为环境问题的重要组成部分, 对人类的身体健康和生态环境造成了严重威胁。为了有效治理重金属污染, 首先需要明确污染源。然而, 由于污染源众多且分布广泛, 传统的监测方法难以准确识别污染源。因此, 建立一种有效的污染源识别模型具有重要的现实意义。

1. 给出主要重金属污染物的分布特征, 并据此识别污染源。
2. 通过数据分析, 说明重金属污染的主要来源。
3. 分析重金属污染物的传播特征, 以此建立模型, 确定污染源的位置。
4. 分析所建模型的有效性, 为重金属污染治理提供科学依据。

### 2 基本假设

1. 各个采样点属于同一污染源, 且数据准确。



- 2) 各区域土壤重金属的背景值相同。不考虑区域功能差异和地形差异。
- 3) 现有重金属的监测结果是污染物正常积累形成的。
- 4) 污染源的位置在局部浓度最高点附近。
- 5) 当监测浓度值大于背景值的1.5倍时, 即认为被污染。

### 3 问题一的模型建立与求解

#### 3.1 8种主要重金属元素在该城区的空间分布

利用39个区域监测点的观测结果, 确定其空间分布, 即要估计区域内任一点 $(x, y)$ 处的浓度值。插值理论可知, 该问题本质是一个插值问题。但, 由于实地人工测量时存在随机性。本题给定的样本观测点在空间上呈现散点式无规律的特征, 显然利用一般的插值或网络插值方法并不可行。为此, 通过文献检索, 本文采用一种基于Dehnay三角剖分的角样条插值方法, 对已知的散点数据进行角样条多项式插值。该方法的数学描述如下<sup>[2]</sup>。

设 $T$ 是三角网格, 具有面 $\tau_i = (x_i, y_i) | i = 1, 2, 3$ 的三角形,  $F_i$ 是面 $\tau_i$ 上 $(x, y)$ 处的观测值(浓度或海拔高度)。构造 $T$ 上该三角形的样条插值曲面函数 $F_T(x, y)$ :

$$F_T(x, y) = W_1(x, y)F_1 + W_2(x, y)F_2 + W_3(x, y)F_3 \quad (1)$$

其中 $W_i(x, y) = W_{j_1}(x, y) = W_{j_2}(x, y)$  是权函数, 且又为

$$W_1(x, y) = \left[ (1/L_1) (15I_1^2 + 6I_1^3 + 30I_1L_1) \left( \frac{Q_1I_1}{D_1} + \frac{Q_2I_2}{D_2} \right) \right] I_1^2$$

$$W_2(x, y) = \left[ (1/L_2) (15I_2^2 + 6I_2^3 + 30I_2L_2) \left( \frac{Q_2I_2}{D_2} + \frac{Q_3I_3}{D_3} \right) \right] I_2^2$$

$$W_3(x, y) = \left[ (1/L_3) (15I_3^2 + 6I_3^3 + 30I_3L_3) \left( \frac{Q_3I_3}{D_3} + \frac{Q_1I_1}{D_1} \right) \right] I_3^2$$

式中 $S$ 为三角形面积

$$L_1(x, y) = \frac{1}{2S} [x_2y_3 - x_3y_2 + (y_2 - y_3)x + (x_3 - x_2)y] \quad D_1 = (x_k - x_j)^2 + (y_k - y_j)^2$$

$$L_2(x, y) = \frac{1}{2S} [x_3y_1 - x_1y_3 + (y_3 - y_1)x + (x_1 - x_3)y] \quad D_2 = (x_i - x_k)^2 + (y_i - y_k)^2$$

$$L_3(x, y) = \frac{1}{2S} [x_1y_2 - x_2y_1 + (y_1 - y_2)x + (x_2 - x_1)y] \quad D_3 = (x_j - x_i)^2 + (y_j - y_i)^2$$

$$Q_1 = (x - x_1)(x_k - x_3) + (y_j - y_1)(y_k - y_3) \quad Q_2 = (x_k - x_j - x_i - x_j) + (y_k - y_j)(y_i - y_j)$$

$$Q_3 = (x_i - x_k)(x_j - x_k) + (y_i - y_k)(y_j - y_k)$$

可以证明 $F_T(x, y)$ 具有如下性质<sup>[2]</sup>:

- a) 在 $T$ 的顶点 $\tau_i = (x_i, y_i)$ 处 $W_i(x, y) = 1$  ( $i = 1, 2, 3$ )
- b)  $W_i(x, y)$ 在 $\tau_i$ 的对边上的函数值为零, 在 $T$ 的边界上, 其一阶法向导数为零
- c)  $W_1(x, y) + W_2(x, y) + W_3(x, y) = 1$

通过给定的采样点数据, 利用上述角样条插值法对样本数据进行插值, 得出拟合后的8种重金属的空间分布, 见图1至图8所示。(左侧为城区等高线图, 右侧为污染物浓度空间分布图。颜色从蓝到红, 由浅到深表示污染物浓度升高)

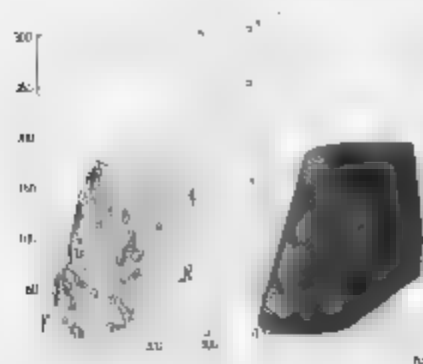


图1: As空间分布

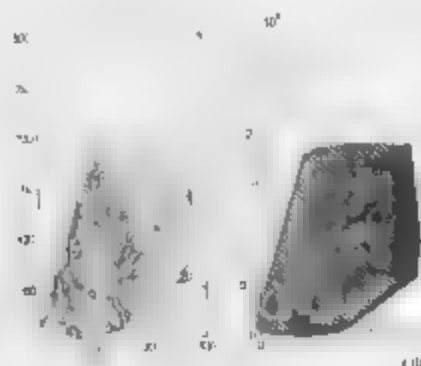


图2: Cd空间分布

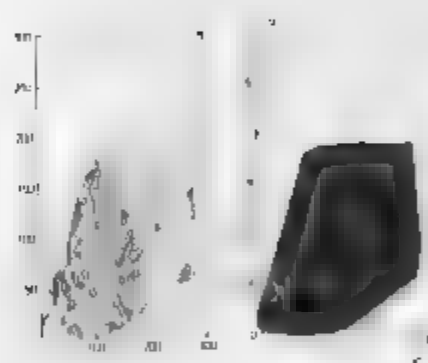


图3: Cr空间分布

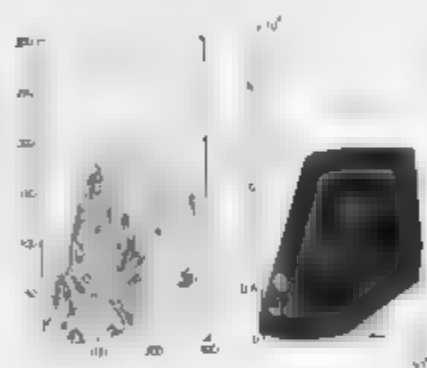


图4: Cu空间分布

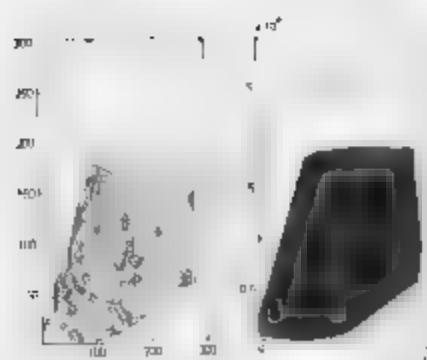


图5: Hg空间分布

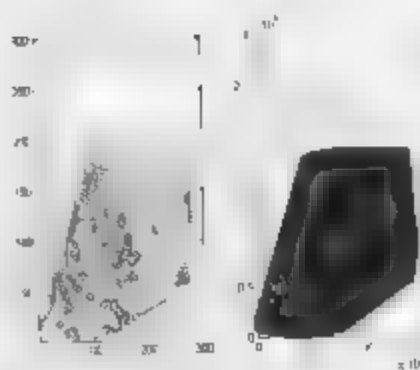


图6: Ni空间分布



### 3.2.2 单因子评价方法

考虑到背景浓度为数值区间, 因此采用类似于 pH 的评价公式<sup>[6]</sup>, 其数学表达式为

$$P_i = \frac{C_i - C_{\min_i}}{C_{\max_i} - C_{\min_i}}, \quad (2)$$

其中  $C_i$  为第  $i$  种污染物的实际监测浓度,  $C_{\max_i}$  为第  $i$  种污染物的背景值上限,  $C_{\min_i}$  为第  $i$  种污染物的背景值下限,  $P_i$  第  $i$  种污染物的污染指数. 如  $P_i < 1$ , 则观测值满足评价标准, 不超标; 否则, 超标, 超标倍数为  $P_i - 1$ .

为反映某区域某污染物的超标范围, 定义超标率. 计算公式如下

$$q = \frac{m}{n} \times 100\%, \quad (3)$$

其中  $m$  为该区域监测结果超标的点位个数,  $n$  为区域中所设的监测点的总数

### 3.2.3 单因子评价结果统计与分析

通过对不同区域各采样的监测结果的单因子评价与结果统计, 得到不同区域各项重金属元素的统计指标, 包括: 采样点数, 超标点数, 超标率, 平均值, 最大超标倍数等. 从统计结果可以看出, 该城区各不同功能区中 8 种重金属元素土壤监测结果均有超标现象, 说明该城区已不同程度地受到人类活动的影响. 其中生活区超标率最高的是 Cr、Cu、Zn、Hg. 最大超标污染物为 Zn 和 Hg, 分别超出背景值标准 49.94 和 19.31 倍. 工业区各重金属元素超标都很严重, 尤其是 Cu、Cd、Pb、Hg、Co, 最大超标污染物为 Hg 和 Co, 分别超出背景值标准 420.28 和 174.17 倍. 因此所有重金属超标率均在 20% 左右, 土壤环境质量相对较好.

道路区除 Cd、Ni 外, 其余超标都很严重, 主要超标污染物为 Cu、Zn、Hg, 最大超标污染物为 Hg 和 Cu, 分别超出背景值标准的 498.41 和 93.36 倍. 公园绿地超标情况也不乐观, 主要超标污染物为 As、Cu、Hg、Zn, 最大超标污染物为 Hg 和 Zn, 分别超出背景值标准 40.26 和 23.08 倍.

### 3.2.4 不同区域污染程度综合评价

1) 评价方法. 对每一个功能区, 采用内梅罗综合污染指数法, 计算公式如下<sup>[7]</sup>

$$P_z = \sqrt{\frac{P_{av}^2 + P_{\max}^2}{2}}, \quad (4)$$

其中  $P_z$  为综合污染指数 (综合反映各污染物对区域土壤的不同作用),  $P_{av}$  为所有单项污染指数的平均值,  $P_{\max}$  为土壤环境中各单项污染指数中的最大值.

2) 污染指数分级方法. 利用参考文献<sup>[4]</sup>, 确定分级方法, 如表 3 所示.

表 3 综合污染指数的分级标准

污染等级	$P_z$	污染程度	污染描述
I	$P_z \leq 0.7$	安全	清洁
II	$0.7 < P_z \leq 1.0$	警戒线	尚清洁
III	$1.0 < P_z \leq 2.0$	轻污染	超标
IV	$2.0 < P_z \leq 3.0$	中污染	土壤 作物受中度污染
V	$P_z > 3.0$	重污染	土壤 作物受严重污染

### 3 评价结果:见表4所示。

表4 内梅罗综合污染指数评价及污染程度分级结果

功能区	1类区	2类区	3类区	4类区	5类区
平均指数值	2.05	5.19	0.69	3.54	1.62
内梅罗综合指数值	3.5	19.48	0.79	13.37	2.02
污染等级	V	V	II	V	IV

4 结果分析 从表4中可以看出,评价城区中,污染最重的区域为2类区(工业区),其次是4类区(道路区),1类区(生活区)和5类区(公园绿地),属轻度污染,3类区、(区)污染最轻。从内梅罗综合指数值来看,总体来说,该城区的重金属污染较为严重。

## 4 问题二的模型建立与求解

首先通过观察图1至图8及表3的数据评价结果,发现主要超标污染物和最大超标污染物具有规律性,这说明不同重金属之间很可能存在很高相关性,即可能来自共同的污染源。为了定量分析各重金属元素之间的关联性,本文利用多元聚类分析对不同重金属进行相关性分析,考察污染物之间的相关性。另外利用一种方法给出不同因子之间的相关系数。若重金属含量有显著的相关性,说明污染源相同,这与污染物产生的原因是密切相关的,因此对照不同污染源的污染特征,可大致确定产生这些污染的原因。最后通过观察地形地貌图,结合海拔高度分布值,给出不同重金属污染的原因。

聚类分析法所采用的相关性度量标准为:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (5)$$

其中 $d(x, y)$ 表示向量 $x, y$ 的距离。

相关性系数法所采用的相关性度量标准为

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (6)$$

通过计算,得各功能区主要元素相关性分析结果,如表5所示。

表5: 各区显著相关或高度相关的重金属

功能区	1类区	2类区	3类区	4类区	5类区
主要相关元素	Cd Pb	Cr-Cu	Cr Ni	Cr-Cu	Cr-Ni
		Cr Hg		Cr Ni	Cu Pb
		Cu Hg		Cu Ni	Pb Zn

按区域分析污染的主要原因



1 各个功能区中,超标率最高的为Cd和Cu,可见该区域是Cd和Zn的高污染区,形成污染的原因有很大可能性来源于生活垃圾填埋或任意排放及含铜农药喷洒所致,也可能来源于相关冶炼企业的烟尘和垃圾排放

2) 生活区中,Cd与Pb显著相关,说明它们污染源相似,可能与生活垃圾中的废电池、电路板及汽车尾气排放有关

3) 工业区内,几乎所有的重金属都显著相关,说明它们污染源相似,而污染程度明显高于其他功能区,可能与工业粉尘和工业垃圾填埋有关,此外Hg的含量明显超标,可能存在含Hg工业垃圾填埋或燃煤电厂粉尘排放

4) 农业区中,各污染物污染指数均较低,Cr与Ni显著相关,说明该区域污染主要来源于生活或工业垃圾污染

5) 交通区中,除Hg之外,其余7种重金属均有显著相关性,说明交通区的Hg来源比较分散,+问题可知,该区域主要超标污染物为Cd、Hg、Zn、Cr、Ni、Pb、Cu,主要来自汽车尾气、轮胎磨损以及大气降尘

6) 饮用水源地中,虽然超标指数较低,高于山区,但是几乎各个指数都超过标准,造成污染的原因可能是因为该区域与受水灌溉或者农业(农药、化肥)造成的

## 5 问题三的模型建立与求解

模型一 基于网格函数拟合的一维样条插值函数逼近污染源识别模型

首先根据题目给出的散点数据,构造基于De Boor 样条函数的角形剖分网格,其次,在剖分网格上分别构造基于浓度和网格高度的二元样条插值多项式(如式(1)所示)

通过角样条函数能够方便地求出任意函数和浓度空间分布函数,通过求导或局部加密网格计算,可以直接得到浓度最高的坐标值及网格高度,即为污染源的坐标<sup>[7]</sup>,以前所未见表6.

表6 污染源位置识别结果

重金属名称	污染源坐标
As	A(18210,13000),B(2650,3023),C(2700,12000),D(4600,7404)
Cd	A(21470,11410),B(4504,11350),C(2300,3085)
Cr	A(3285,5862)
Cu	A(2300,3517)
Hg	A(2683,2406),B(5240,9134),C(13800,2345)
Ni	A(3354,5862)
Pb	A(2108,3332),B(2792,4996)
Zn	A(13800,9626),B(9487,4443)

模型二 对流-扩散方程模型

雨水的作一是重金属在土壤表面传播的主要途径之一,重金属在土壤中主要以溶质的形式存在,而其传播方式主要以对流扩散方式为主,由此建立对流扩散方程<sup>[8]</sup>

$$\frac{\partial C}{\partial t} + D \frac{\partial^2 C}{\partial x^2} = u \frac{\partial C}{\partial x} \quad (7)$$

其中  $\frac{\partial C}{\partial t}$  为浓度关于时间的导数,  $D \frac{\partial^2 C}{\partial x^2}$  为扩散项 (表明自然扩散的速率正比于浓度的梯度),  $D$  为弥散系数,  $u \frac{\partial C}{\partial x}$  为对流作用产生的污染物浓度的迁移速率,  $u$  为对流速度。在本文所研究的问题中, 对流速度  $u$  可视为与地形高度梯度成正比的函数, 即  $u = k \nabla h(x, y)$ ,  $h(x, y)$  为地形高度曲面函数 (可由模型一近似得到)。

在给定初始和边界条件下, 式(7)的求解可求解析法或数值计算方法, 模型解。但在实际处理中, 难以准确地确定边界条件, 因此直接求解模型存在实际的困难。

### 模型三 分布模型

在文献[6]中, 标准的对流扩散方程存在解析解表达式。它的解为正态分布函数。随着时间变化, 浓度趋于均匀分布。另一方面, 与波动方程类似, 自两个方向传播。本质上是一个正态分布模型。

#### 正态分布模型 (对流为主)

在地理现实中, 污染物在平面方向上的分布是不均匀的。一般情况下, 沿河流平缓的方向的分布是近似理想的正态分布。而如果两边出现障碍, 可能会使分布更加向中心集中。因此, 用高斯分布的一维分布模型拟合, 可以达到更准确的效果。

$$C(x, y) = C_0 \exp \left[ -\frac{\lambda_1 (x - x_0 + b_2 y - y_0)^2 + \lambda_2 (x - x_0 + b_2 y - y_0)^2}{2} \right] \quad (8)$$

其中  $\lambda_1, \lambda_2, a_1, a_2, b, b_2$  为待定参数,  $x_0, y_0$  为污染源位置坐标,  $C_0$  为污染源强度,  $C(x, y)$  为污染物在  $(x, y)$  处的浓度值。

#### 均匀分布模型 (扩散为主)

如果污染物在地理各方向的传播都受到限制, 则污染物会在污染源附近堆积, 这时可以将污染物的分布特征用均匀分布近似地表示出来。

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f \quad (9)$$

若  $F(x, y) < 0$ ,  $F(x, y) = C_0 - F(x, y) > 0$ ,  $C(x, y) = C_1$ , 其中  $C_1$  为污染源区域内部的点的浓度,  $C_0$  为污染源外部的点的浓度。

#### 模型求解方法

- 1) 选取疑似污染源附近网格作为局部模拟区域, 得到局部插值函数。
- 2) 加密局部网格, 计算插值函数, 计算各网格点的函数值, 作为观测值。
- 3) 用线性最小二乘法建立解析解与观测值的误差函数, 求解待定参数, 使误差平方和最小。

$$\min f(a, x_0, y_0) = \sum (C(x_i, y_i) - C(x_i, y_i, a))^2 \quad (10)$$

其中  $a$  为待识别参数向量。

- 4) 识别污染源位置。所求的  $(x_0, y_0)$  即为该污染源的中心位置。再由地形曲面插值函数, 计算空间海拔高度, 即「确定污染源的实际情况」。

通过模拟、仿真发现, 求解结果与模型一基本相一致。

## 6 问题四的模型建立与求解 (略)

## 参考文献

- 孙道生, 陈玉娟, 阮国标. 城市及近郊土壤重金属含量特征及人类活动的影响[J]. 中国环境科学, 2001, 40(4): 93-96.

- Guan D S, Chen Y J, Ruan C B. Study on heavy metal concentrations and the impact of human activity on them in urban and suburb soils of Guangzhou[J]. ACTA Scientiarum Naturalium Universitatis Sunyatseni, 2001, 40(4): 93-96
- [2] 张基明, 孙德汉. 工业 散乱数据点的三次多项式插值[J]. 计算机辅助设计, 工程图学学报, 1998, 10(5): 416-424
- Zhang C M, Sun D F, Wang J Y. Cubic polynomial interpolation of scattered data points[J]. Journal of Computer-Aided Design & Computer Graphics, 1998, 10(5): 416-424
- [3] 蔡心雄, 李基.  $C^1$ 自然插值逼近曲面拟合[J]. 数学工程, 2010, 1: 1-10
- Nie Z F, et al. Surface fitting based on  $C^1$  natural neighbor interpolation[J]. Journal of Engineering Graphics, 2010, 1: 110-115
- [4] 董少萍, 杨俊衡. 城市展点土壤[4]土壤重金属污染评价指标探讨——以贵阳市某区为例[J]. 安全与环境工程, 2010, 17(1): 57-64
- Jiang Z P, Yang J H. Discussion on heavy metal quality assessment standards of urban pollution monitoring (taking Hongyang city for example) [J]. Safety and Environmental Engineering, 2010, 17(1): 57-64
- [5] 夏林, 王洪. 土壤污染及其防治[M]. 上海: 同济大学出版社, 2001
- Xia L J, Wang H K. Soil Pollution and Prevention[M]. Shanghai: East China University of Science and Technology Press, 2001
- [6] 余剑东, 倪玉珍, 杨开明. 土壤重金属污染评价指标的研究进展[J]. 安徽农业科学, 2002, 40(5): 1-7
- Yu J D, Ni W Z, Yang X T. Current progress in evaluation indexes for heavy metals pollution in soils[J]. Guangdong Trace Elements Science, 2002, 9(5): 11-17
- [7] 周博. 用 Matlab 科学计算[M]. 北京: 机械工业出版社, 2010
- Zhou B, et al. Scientific Computing with Matlab[M]. Beijing: China Machine Press, 2010
- [8] 宋学森, 吴玉玲, 陈迪. 基于 Matlab 的挥发源项方程的三角插值解[J]. 环境科学与工程, 2008, 31(11): 31-34
- Song X S, Wu Y L, Chen D. Numerical simulation for volatile source terms in convection-dispersion equation based on Matlab[J]. Environmental Science & Technology, 2008, 31(11): 31-34

## Study on Urban Heavy Metal Pollution Assessment and Source Recognition Model Assessment

QI Qian, CHENG Sheng-hui, YAN Bin-bin

Advisor: LIU Bao-dong

(Shandong University, Jinan 250100)

**Abstract:** According to the scattered features of the given data, a triangle spline interpolation method is constructed based on the DeSoney triangular subdivision technique. By utilizing this method, the spatial distribution for each heavy metal is then obtained. Based on the fact that the background value corresponds to an interval for each heavy metal, the factor assessment model and the comprehensive evaluation method are established. The assessment result for each subregion is also provided. Besides, by using the multivariate clustering and correlation methods, the relationships between different heavy metals is obtained and the main reason underlying the urban heavy metal pollution is evaluated. To recognize the source location of the heavy metal pollution, three models are constructed respectively. The first one is established based on the triangular spline interpolation technique, the second based on the characteristic of convection-dispersion, and the third based on the distribution characteristics of convection or dispersion. By combining these three models, the method for recognizing the source location is proposed and its effectiveness is substantiated by experiments.

**Keywords:** triangle spline interpolation; clustering analysis; correlation analysis; convection-dispersion equation; normal distribution fitting; evenly distributed fitting



表1: Muller地累积指数污染物等级划分

$I_{geo}$ 指数	< 0	0 ~ 1	1 ~ 2	2 ~ 3	3 ~ 4	4 ~ 5	> 5
指数	0	1	2	3	4	5	6
累积污染程度	无	无中	中	中强	强	强极强	极强

1) 处理的数值 $\times$ 城并不是现成的地形。本文利用ArcGIS 9.3 把1km分辨率的观测值利用克里金插值插入到73m分辨率的栅格网络中。对每个栅格 $\times$ 城有按高度、功能 $\times$ 以及各中金属污染等级分布图,浓度指数分布图。以Hg为例,如图1和图2。



图1: Hg的污染分级图



图2: Hg的污染指数图

综合考虑城 $\times$ 地中以及各个元素的污染分级和污染指数后,对8种元素浓度值的空间分布做讨论后,得到表2。

表2: 污染指数分布的大致趋势

## 重金属 地累积指数 污染分布

As	(0.5015)	东部和中部无污染 西南地区轻度 中强污染
Cr	(0.816)	东部和中部无污染 西南地区轻度 中强污染到中度污染
Au	(0.806)	东部和中部无污染 西部轻度 中强污染 但污染区域均匀且范围较大
Cd	(0.412)	东部无污染 中部和西部轻度 中强污染 西南角中度污染
Pb	(0.618)	东部无污染 中部和西部轻度 中强污染 西南角中度污染
Zn	(0.822)	东部无污染 中部轻度 中强污染 西部中度污染 西南角中强度污染
Cu	(0.225)	东北角无污染 东部和中部轻度 中强污染 西部中度污染 西南角中强度污染
Hg	(2.5)	东部无污染 中部有数值的轻度 中强污染 西部中度污染 呈多峰值散乱分布

## 4.2 对每个功能区的污染程度分析

地累积指数分析了每种重金属的污染分布,但不同污染物、污染源有不同的环境效应、不同的危害,为了评定不同功能区的重金属的综合污染程度 使之能在同一尺度上比较,引入的



染负荷指数的分析各功能区的污染程度

$$C = \frac{C_i}{C_{0i}} \quad PII = \sqrt[n]{\prod_{i=1}^n C} \quad PII_{zone} = \sqrt[m]{\prod_{j=1}^m PII_j} \quad (2)$$

式中  $C_i$  为第  $i$  种金属浓度的实测值,  $C_{0i}$  为第  $i$  种金属的背景值,  $PII$  为单个点的污染负荷指数,  $n$  为金属污染物种数,  $PII_{zone}$  为  $k$  区域的污染负荷指数,  $m$  为采样点个数。根据  $PII$  的值, 规定了相应的污染负荷指数等级划分标准, 见表 3。

表 3: 污染负荷指数等级划分

$PII$ 值	$\leq 1$	1-2	2-3	$> 3$
污染等级	0	1	2	3
污染程度	无污染	中等污染	强污染	极强污染

利用式 (2) 计算得到: 工业  $\times$ 、1 号道  $\times$ 、生活区  $\times$ 、公园绿地  $\times$ 、1 区的  $PII_{zone}$  分别为 2.15、1.02、0.83、0.81、0.6。即五大功能区的污染程度为: 工业  $\times > 1$  号道  $\times > 生活区 \times > 公园绿地 \times > 1$  区。这 5 个符合生活区及 1 号道  $\times$  污染最强, 其余四个功能区污染程度都为 1 等污染。在 1 区及 1 号道更核心处污染。

## 5 对问题二的分析和处理

本文的假设所给的金属污染模式相同, 因此若重金属除了具有类似的分布, 可以认为它们具有共同的来源。根据问题一, 对每种元素污染指数分布的分析, 将分布相同或类似的元素合并成 4 个类型, 如表 4。

表 4: 每一类的典型重金属污染分布特征

类型	元素	特征
1	Ni, Cr	污染程度最低, 东部和中都没有污染, 西部少量地区有轻微污染。污染范围小。
2	As	污染程度较低, 东部和中都无污染, 西部轻度、中度污染, 但污染区域均匀且范围较大。
3	Cd, Pb, Zn, Cu	污染程度中等偏高, 污染等级呈阶梯分布, 受中等污染的范围较大。个别地区出现强污染。
4	Hg	污染程度较高, 污染分布散乱, 但污染强度总体污染程度较低, 但在某些点出现极强污染。

下面用统计方法<sup>[24]</sup>分别针对标准化浓度和地累积指数验证表 4 分类的合理性。标准化数据浓度是对原始浓度标准化后得到的, 它能反映重金属浓度空间分布的实际情况。地累积指数则是考虑到了重金属浓度的背景值信息, 侧重于评定重金属污染等级。因此这两种数据类型各有侧重。

利用 SPSS 18.0, 计算 8 种元素的地累积指数的简单相关系数矩阵, 得到 Zn、Pb、Cd、Cu 之间存在非常人的相关性, Cr 和 Ni 的相关性比较大, As 和 Hg 是两种比较独立的元素, 与其他几种因素指数的空间分布相关性不是非常显著。

对标准化浓度数据的相关分析可得, Cr 和 Ni 之间存在显著的相关性, Pb 和 Cd、Pb 和 Cu, 以及 Pb 和 Zn 之间相关性也比较显著。而 As 和 Hg 与其他元素的相关性相对较小, 属两种比较独立的元素。地累积指数和标准化浓度数据相关性的结论与表 4 的结论相符合, 说明了地累积指数在该问题上能够较好的反映区域的污染情况。另外也证明了表 4 分类的合理性。

利用 SPSS 18.0 对 8 种重金属空间分布在表 1 上成分分析法分析, 聚成因。结果表明 8 种重金属 Cd、Cr、Cu、Ni、Pb、Zn, 能有一个共同的来源。As 和 Hg 比较独立。

综上所述可得, Ni 和 Cr 相关性最显著, Pb、Zn、Cd、Cu 的相关关系显著。我们假定它们分为独立的两类, 并, 每一类具有单一污染来源。而 As 和 Hg 为比较独立的两种元素, 因此认为它们源自不同的污染源。从《中国土壤重金属来源》可以总结出 Pb 主要源自工业、废气、Cu 主要源自工业、农业、Ni、Cr、Cd、Zn 主要源自工业。As 主要源自矿业、采和农业, Hg 的主要源自废水和生活垃圾。为了全面的分析题目所给区域污染形势, 提炼了污染分布 5 个关键区。即污染高值区。如图 3, 关键区中包含 2 个浓度异常点(A 点和 C 点)对这些区域进行详细分析。进一步确定了污染最可能的来源。

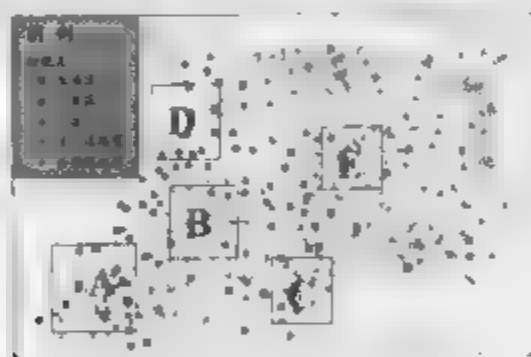


图 3 污染集中分布区域图

例如在 B 区, Pb、Cr、Cu、Zn 都有高污染值。根据国家根据《中国土壤重金属来源》, 工业废水、废水、废气都是这 4 种金属的来源, 除此外 Cu 还可以来源于农业, Pb 还可以来源于汽车尾气。但从图中可以看出 B 区主要是道路交通主干道, 因此我们猜测该区域 Pb 的一个来源可能是汽车尾气。而 B 区主要处于人类的工业区, 但这 4 种重金属污染反而不大, 我们进一步猜测在 B 区有一集中的垃圾场, 在此, 集中的工业废物是 4 种金属的其他来源。对其余几个关键区类似分析在此不再赘述。

## 6 对问题三的分析 and 处理

### 6.1 扩散模型

设有一污染源, 沿  $x, y, z$  个方向的扩散, 扩散系数分别为  $a^2, b^2, c^2$ , 设  $u(x, y, z, t)$  是  $t$  时刻点  $(x, y, z)$  处某物质的浓度。任取一个闭曲面  $S$ , 围成的区域为  $\Omega$ , 由扩散理论中的费克第

定律,  $\Delta t$  时间内, 通过  $S$  流入  $\Omega$  的质量为

$$M_1 = \int_t^{t+\Delta t} \iint_S \left( a^2 \frac{\partial u}{\partial x} \cos \alpha + b^2 \frac{\partial u}{\partial y} \cos \beta + c^2 \frac{\partial u}{\partial z} \cos \gamma \right) dS dt. \quad (3)$$

由于衰减,  $\Delta t$  时间内  $\Omega$  中的质量减少为

$$M_2 = \int_t^{t+\Delta t} \iiint_{\Omega} k^2 u dV dt, \quad (4)$$

其  $k^2$  为衰减系数<sup>[5]</sup>,  $\Delta t$  时间内  $\Omega$  内浓度变化引起的质量增加为

$$M_3 = \iiint_{\Omega} [u(t + \Delta t) - u(t)] dV = \int_t^{t+\Delta t} \iiint_{\Omega} \frac{\partial u}{\partial t} dV dt \quad (5)$$

由物质质量守恒定律, 显然  $M_3 = M_1 - M_2$ , 于是得衰减的扩散过程的数学模型为

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + b^2 \frac{\partial^2 u}{\partial y^2} + c^2 \frac{\partial^2 u}{\partial z^2} - k^2 u \quad (6)$$

设扩散源在  $C_0(x_0, y_0, z_0)$  处, 则此扩散问题满足 Cauchy 问题. 用傅立叶变换可求得解析解为

$$u(x, y, z, t) = \frac{M}{8\pi tabr\sqrt{\pi t}} \exp \left\{ -\frac{(x-x_0)^2}{4a^2 t} - \frac{(y-y_0)^2}{4b^2 t} - \frac{(z-z_0)^2}{4c^2 t} - k^2 t \right\}, \quad (7)$$

式中  $M$  为扩散源的质量.

在无限大介质中扩散浓度集中在点源, 这类扩散问题都具有高斯分布的形式, 即“高斯解”. 此受到启发, 假设污染扩散浓度模式为多元的正态分布曲面. 多元正态分布是多元正态分布的直接推广. 设随机向量  $X = (x_1, x_2, \dots, x_p)^T$  服从  $p$  维正态分布, 则其密度函数为

$$f(X) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu) \right\}, \quad (8)$$

或可简化地表示为

$$g(u) = k \exp \left( -\frac{u}{2} \right), \quad (9)$$

其中  $k$  为污染源的权重系数,  $u$  为马氏距离.

题目所给的浓度信息是包含高度信息的空间数据, 为简化求解先把题目所给的空间数据通过某种方式投影到平面上, 然后在水平面上模拟污染分布.

## 6.2 高度订正方法

第  $i$  个观测点的污染物浓度指数可表示为

$$I_i = I(x_i, y_i, h_i), \quad i = 1, 2, \dots, n, \quad (10)$$

其中  $x_i, y_i$  是观测点的水平位置,  $h_i$  是观测点的海拔高度. 将位置函数  $f$  与海拔高度  $h$  做变量分离后得到

$$I_i = \tilde{f}(x_i, y_i) J(h_i) + \varepsilon_i, \quad (11)$$

其中  $\varepsilon_i$  是随机误差, 服从正态分布,  $\tilde{f}(x_i, y_i)$  是浓度随水平位置的变化,  $J(h_i)$  是浓度随高度扩散函数. 该式表明第  $i$  个观测点的污染物浓度可以近似等效为水平扩散和垂直扩散的叠加作用造成.

为得到污染浓度与地形高度之间的关系, 必须消除污染指数随水平传播距离变化的影响。选取城区西南角的一处孤立且陡峭的山坡拟合污染传播高度为自函数, 该处重金属浓度受高度的影响远远大于受水平距离的影响。

根据比例-伯格-朗伯定律, 假设重金属污染扩散随高度呈指数衰减, 于是用指数函数在上述区域对污染传播高度为自函数进行估计。将地形指数做高度订正, 从而将原始样本数据  $(x_i, y_i, h_i, I_i)$  换算成新样本  $(x_i, y_i, I_i)$ , 从而为接下来的位置估计降维。

### 6.3 模型求解和污染源的搜索

当只有一个污染源时,  $(x, y)$  处的浓度分布为

$$\tilde{I}_i = k \exp\left(-\frac{u_i}{2}\right), \quad (12)$$

当有  $N$  个污染源时,  $(x, y)$  处的浓度是  $N$  个污染源在此处污染效果的叠加

$$\tilde{I}_i = \sum_{j=1}^N k_j \exp\left(-\frac{u_{ij}}{2}\right) \quad u_{ij} = (\mu_j - \mu_0')^T \Sigma_j^{-1} (\mu_j - \mu_0')$$

设  $\mu_j = x_j, \mu_0' = x_0, \mu_0 = x_0$  这里  $k, \Sigma_j, \mu_0' (j=1, 2, \dots, N)$  是待估参数

然后, 通过似然估计或最小-乘法进行污染源搜索, 即可估计出估参参数  $\mu_0, \mu_0', k, \Sigma_j, N$ , 从而为污染源位置。这里, 污染源的个数可依据 AIC 模型选择标准或交叉验证法选取。具体的搜索步骤如下:

**步骤 1** 利用公式 (12) 将高度为  $h_i$  的浓度根据指数衰减规律订正到了平面  $h_0$ , 得到订正后的重金属污染等值线图;

**步骤 2** 分析  $h_0$  重金属污染指标的等值线图, 找到可能的污染源位置;

**步骤 3** 找出步骤 2 中污染区内重金属污染指标的极大值点, 作为污染源位置参数  $\mu_0'$  的初始位置, 搜索区域为污染源周围一定领域;

**步骤 4** 进行区域搜索。根据最小-乘法找出对应的污染源位置;

**步骤 5** 根据 AIC 准则确定最佳的污染源的个数及其位置;

下面以第一类元素 (包含 Cr 和 Ni) 为例, 选取 Cr 作分析, 给出具体的操作过程。根据高度  $h_0$ , 污染分布图 1, Cr 能仅有一个单污染源形成或者一个双污染源形成。



图 4 第一类金属元素的污染分布及污染源搜索

若为单污染源。搜索区域选为  $X \sim [1000, 9000], Y \sim [1000, 9000]$ , 找到污染源的位置为 (4592, 4604)。

若为双污染源,搜索区域选为  $X \sim [1000, 12000]$ ,  $Y \sim [1000, 9000]$ , 初值取该搜索区域内污染物含量指数的极大值所在位置  $\mu_0(4777, 4897)$ ,  $\mu_1(10643, 4772)$ , 计算后找到污染源的位置为  $\mu'_0(4586, 4620)$ ,  $\mu'_1(15291, 3208)$ 。

分析上面的搜索结果,可知双污染源得到的两个污染源的位置中,  $\mu'_1(15291, 3208)$  位于 L<sub>1</sub> 区,而在污染指标等值线分析图中是表现为一个低值,认为这种分析结果并不合理,对此我们认为污染源应为一个单污染源,因此第一类元素的污染是由位于 (4592, 4604) 的单污染源形成的。

## 7 总结

该模型所分析的对象都是重金属浓度的地累积指标,而不是直接使用测到的浓度,从而消除了当地环境对重金属元素的分布影响,根据污染等级的划分,分析出整个 L<sub>1</sub> 区都不属于污染区。在污染源考虑时略去了侧面的 L<sub>2</sub> 区,大大简化了问题。测量值是空间数据,包括高程信息。我们在分析污染前对高程先进行打点,以降低模型维数,直接增大样本容量。衰减模型利用北方吸收原理的衰减原理去拟合样本,得到令人满意的效果。结合污染源扩散污染物浓度的椭圆型分布规律,由分布的位置参数确定污染源位置,并且可以处理多个污染源的污染叠加效应。

1) 数据信息限制,模型只考虑了点污染源分布,没有考虑线污染源和其它形式污染源的扩散分布。如果增加信息,则可进一步改进模型。

2) 继续采样,增加样本信息,得到更为真头的重金属污染物浓度指标分布  $C(x, y, z_1, z_2)$ 。

3) 增加城市的地表信息,如河流、农田等的分布状况,可以估计一些非点源分布的污染源。

4) 考虑风同等引起的污染分布的不对称性,将污染浓度曲面从二维态模型  $g(x)$  扩展至偏斜二维分布  $2g(x, G(\lambda, \mu))$ , 其中  $G(\cdot)$  为一元正态分布函数,  $\lambda$  为偏斜参数。

## 参考文献:

- 1 朱卫华. 地累积指数法在土壤重金属污染评价中的应用[J]. 同济大学学报(自然科学版), 2006, 34(12): 1657-1661.
- 2 Cha S W. Application of index of geoaccumulation to pollution evaluation of heavy metals in soil[J]. Journal of Tongji University (Natural Science), 2006, 34(12): 1657-1661.
- 3 李向东. 应用多元统计[M]. 北京: 北京大学出版社, 2008.
- 4 Li W D. Applied Multivariate Statistical Analysis[M]. Beijing: Peking University Press, 2008.
- 5 林艳. 基于地统计学与GIS的土壤重金属污染评价与预测[D]. 长沙: 中南大学硕士学位论文, 2009.
- 6 Lin Y. Pollution assessment and prediction of heavy metals in soil based on geostatistics and GIS[D]. Changsha: Master Thesis of Central South University, 2009.
- 7 徐明超. 长沙市农田土壤重金属污染的空间分布及污染评价[D]. 重庆: 西南大学硕士学位论文, 2007.
- 8 Xu L C. Spatial pattern and pollution evaluation of heavy metals in farmland soils of Fuxin city[D]. Chongqing: Master Thesis of Southwest University, 2007.
- 9 李庆波. 选煤厂生产环境粉尘扩散力模型建立[J]. 北京: 北京职业技术学院学报, 2008, 7(1): 84-89.
- 10 Li Y Q. Construction of diffusion equation in the process of engendering dust in selecting coal factory[J]. Journal of Beijing Polytechnic College, 2008, 7(1): 84-89.
- 11 陈杰. Matlab 宝典(第3版)[M]. 北京: 电子工业出版社, 2011.
- 12 Chen J. Matlab Book, 3rd Edition[M]. Beijing: Electronics Industry Press, 2011.
- 13 滕波. 重金属污染对农产品的危害与风险评估[M]. 北京: 化学工业出版社, 2010.
- 14 Teng W. Heavy Metals Pollution on Agricultural Products Hazard and Risk Assessment[M]. Beijing: Chemical Industry Press, 2010.



## Spatial Data Analysis of Heavy Metal Pollution in Urban Soil

SHENG Zhen-feng, HUANG Si, ZHOU Min-qiang  
Advisor: CAO Chun-zheng

(Nanjing University of Information Science & Technology, Nanjing 210044)

**Abstract.** With the rapid development of current industry, heavy metal pollutants (HMP) have made a tremendous impact on the environment. The degree of HMP is measured by the geoaccumulation index (GI), and meanwhile, the degree of HMP in each functional areas is assessed by the pollution load index (PLI). After such analysis, a categorization of heavy metal ions is concluded. Based on the national standards for the categorization of HMP's sources and the distributed characteristics of the pollution, the pollution source is inferred for each HMP's ion. Following the revision of the sample's elevation information through the exponential smoothing, the pollutant concentration surfaces formed by multi-sources are fitted by the density of weighted mixture of bivariate normal distributions. The estimates of the location parameters can be regarded as the HMP's sources.  
**Keywords:** heavy metal pollution analysis; geostatistical analysis; spatial data; geoaccumulation index; mixture of normal distributions



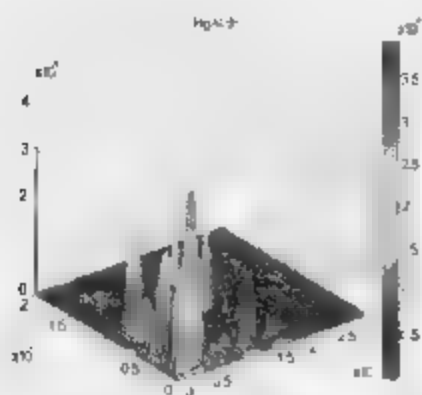


图1 重金属Hg在流域口的空间分布

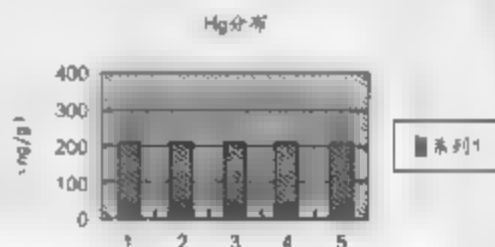


图2 重金属Hg在流域口的海拔分布

## 2 问题一(2)模型建立与求解

主分量分析评价法 题目要求我们分析流域内不同城市区域的污染程度。某流域内不同城市区域、该流域各种水功能区、案情以及其他因素。本文旨在通过主分量分析、因子分析、除一般取单因子、主因子、主成分、从而更科学、具体地解决如下问题。

### 1) 数据标准化处理

1.1 各参数的量纲、大小不同,首先应将其各变量数值标准化,统一比较的尺度。用一分指数公式将参数数据标准化。

$$x_{ij} = (C_{ij} - S_j) / (S_j - B_j), \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n,$$

其中 $x_{ij}$ 代表第 $i$ 个样本第 $j$ 种参数的标准化值, $C_{ij}$ 是第 $i$ 个样本第 $j$ 种参数的 $N$ 个原始观测值的平均值, $S_j$ 代表第 $j$ 种参数的最高观测值(背景值), $B_j$ 代表第 $j$ 种参数的最低观测值。 $m$ 代表样本个数, $n$ 代表参数个数。

### 2) 主分量分析法

将 $m$ 个样本的 $n$ 个参数数据经过标准化处理,构成一个标准化数据矩阵 $X_{n \times m}$ 。用Matlab求 $X$ 的协方差阵 $C = \frac{1}{m} X X^T$ 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 及相应的特征向量 $v_1, v_2, \dots, v_n$ 。将特征向量归一化,得到正交阵 $V$ ,作变换 $Y = V^T X$ ,特征值 $\lambda$ 就是新变量 $Y$ 的方差。将 $\lambda$ 个特征值按大小顺序排列 $\lambda_1, \lambda_2, \dots, \lambda_n$ ,相应的 $n$ 个特征向量组成 $n$ 个新变量 $y_1, y_2, \dots, y_n$ 。新的变量对模型贡献较大。新变量 $y_1, y_2, \dots$ 分别称为第1主分量,第2主分量,……。前面几个主分量构成了样本间的最大变化特征:

$$\begin{cases} y_1 = v_1^{(1)} x_1 + v_2^{(1)} x_2 + \dots + v_n^{(1)} x_n, \\ y_2 = v_1^{(2)} x_1 + v_2^{(2)} x_2 + \dots + v_n^{(2)} x_n, \\ \vdots = \dots + \dots + \dots + \dots, \\ y_m = v_1^{(m)} x_1 + v_2^{(m)} x_2 + \dots + v_n^{(m)} x_n \end{cases}$$

前  $p$  个主分量  $y_1, y_2, \dots, y_p, p < n$  的方差占总方差的比例为:

$$\rho = \sum_{i=1}^p \lambda_i / \sum_{i=1}^n \lambda_i$$

$\rho > 0.7$  时, 前  $p$  个主分量解释  $\lambda_i$  占  $70\%$  以上, 此时可选取前  $p$  个主分量转换来  $n$  个变量,  $n$  个变量在第  $i$  个公共因子上的荷载向量为

$$a_i = \sqrt{\lambda_i} v^i = \sqrt{\lambda_i} (v_1^{(i)}, v_2^{(i)}, \dots, v_n^{(i)})^T, \quad i = 1, 2, \dots, p$$

第  $j$  个变量在全部  $p$  个公共因子上的荷载的平方和叫做变量  $x_j$  的公共性

$$h_j^2 = \sum_{i=1}^p a_{ij}^2, \quad j = 1, 2, \dots, n$$

从而有

$$h_j = \sqrt{\sum_{i=1}^p a_{ij}^2}$$

的  $\lambda_i$  按照变量  $x_j$  在公共因子上的荷载重要性程度, 把每个变量  $x_j$  的公共性的平方根  $h_j$  作为变量  $x_j$  的权重, 则构成一个度量环境品质优劣的综合评价指标

$$Y = \sum_{j=1}^n h_j x_j$$

式 (1) 中  $Y$  为指数表示的某样本中第  $j$  种参数的标准值,  $i$  在数据标准化处理时所占权重, 即  $Y$  为第  $j$  种参数的权重 (指数) 的表示,  $x_j$  为某样本中环境品质优劣的综合评价指标,  $h_j$  为评价指标的权重进行环境品质评价。

### 3) 模型实现

本文将 8 种不同重金属的浓度在 313 个采样点的数值和背景土壤值 (28 组数据) 带入模型, 将背景值与最大点值作为最大点值进行归一化处理。再在 Matlab 中编写程序得到 313 个采样点的背景值的综合污染指数。最后将背景值与最大点值进行归一化处理, 得到 313 个采样点的综合污染指数。再用 Kriging 插值, 对综合污染指数进行拟合, 得到综合污染指数的空间分布和等值线分布 (如图 3 和图 4), 其中横坐标为  $x$ , 纵坐标为  $y$ , 横坐标一个单位代表 300, 纵坐标一个单位代表 200。

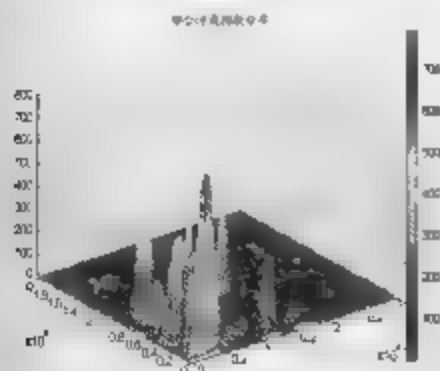


图3 综合污染指数的空间分布



图4 综合污染指数的等值线分布

对各个功能区污染程度统计,如表1.

将整个城区按横纵坐标等分成 $5 \times 5$ 的网格,各 $\times$ 域内采样点的综合采样指数做平均后得到各个区域的污染情况统计,见表2.0表示该 $\times$ 域没有采样点.

表1: 各个功能区污染程度统计表

区域	As( $\mu\text{g/g}$ )	cd( $\mu\text{g/g}$ )	Cr( $\mu\text{g/g}$ )	Cu( $\mu\text{g/g}$ )	Hg( $\mu\text{g/g}$ )	Ni( $\mu\text{g/g}$ )	Pb( $\mu\text{g/g}$ )	Zn( $\mu\text{g/g}$ )	综合指标
生活区	5.5	295.7	52.8	47.1	208.5	17.1	59.8	135.7	5.0
工业区	5.7	311.9	55.9	60.6	339.0	17.7	64.6	217.8	7.2
公园	5.3	26.8	41.4	34.1	207.0	16.2	52.2	135.8	4.7
主干道路 $\times$	5.7	302.9	53.5	55.1	300.5	17.2	61.8	211.5	6.4
公园绿地 $\times$	5.6	295.0	52.6	46.1	202.8	17.4	59.1	192.2	4.9

表2: 各个区域的平均污染情况统计表

	[0, 4000)	[4000, 8000)	[8000, 12000)	[12000, 16000)	[16000, 20000)
[0, 6000)	24.17	7.92	3.88	0.00	0.00
[6000, 12000)	3.11	6.63	4.65	5.56	1.32
[12000, 18000)	20.14	1.37	12.90	2.40	1.44
[18000, 24000)	1.00	1.20	2.80	1.03	1.14
[24000, 30000)	0.00	1.62	0.99	0.97	1.06

由以上两表格,对五种功能区,污染情况从重至轻依次是:主干道路 $\times$ 、工业区、生活区、公园、公园绿地 $\times$ . $\times$ 域在城市内有一块 $\times$ 域污染较为严重,污染源很可能位于其上, $[0, 6000) \times [0, 4000)$ 、 $[12000, 18000) \times [0, 4000)$ 、 $[12000, 18000) \times [8000, 12000)$ .

### 3 问题二模型建立与求解

题目要求通过数据分析,说明重金属污染的主要原因.本文总结城市土壤重金属污染的七大成因,即生活垃圾、工业生产、自然堆积、交通污染、植物吸附、人口密度和化肥农药.采用层次分析法,将总污染指数作为最上层目标层,五个功能区作为中层准则层,七种污染原因作为最下层方案层.通过问题一中各个功能区污染情况的统计数据,确定准则层到目标层的权重.为各个污染原因对各个功能区污染的贡献赋值,获得方案层对准则层的权重,最终获得方案层到目标层的权重.即各个污染因素对污染的影响.

首先概述层次分析法<sup>[2]</sup>层次分析法(The analytic hierarchy process)简称AHP,它是一种定性和定量相结合的、系统化、层次化的分析方法.一般包含4个步骤:1)建立层次结构模型;2)构造两两比较矩阵;3)计算权重向量并做一致性检验;4)计算组合权重向量并做组合一致性检验.下面借助Excel与Matlab进行上述步骤.

### 1) 建立层次结构模型

对于本问题,目标层的任务是决定七人因素中哪些因素是该城区重金属污染的主要成因。准则层是生活区、工业区、山区、主干道区、公园绿地。方案层为生活垃圾、工业生产、自然沉降、交通污染、植物吸附、人口密度和化肥农药七人因素。

通过查阅相关资料<sup>[4]</sup>,得知可能导致污染的原因如下:

+生活垃圾:废弃的生活物品中含有重金属元素,比如电池;工业生产:化学实验产生含重金属的废水废渣,自然沉降:描述没有任何生物作用的情况下重金属在自然界的分布与转化;交通污染:车辆排放含有重金属的尾气;植物吸附:某些植物的根可以吸附土壤中的重金属形成聚集;人口密度:其他人为活动对重金属含量的影响,如化学实验、化肥农药。化肥农药在土壤中仍可能有残留。

针对问题(1)的七人类生活区域,它们各自有不同的影响对象,见图5。生活区:人类生活垃圾,存在交通污染情况;人口密度大。工业区:进行工业生产的区域,附近交通运输频繁。人口密度稍小。山区:山区人口密度较小,存在植物的吸附作用。主干道区:交通污染程度严重。公园绿地:交通污染程度轻,植物主要集中在公园绿地,撒播的农药有残留。自然沉降普遍存在于各个区域。

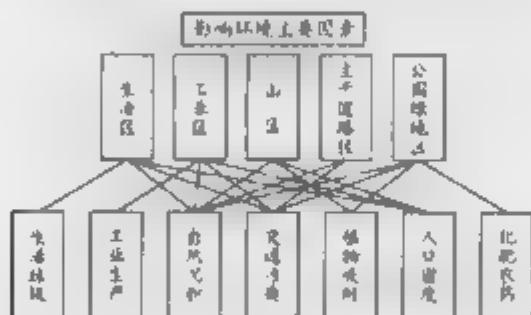


图5: 五大类区域不同的污染因素

表3是准则层与方案层的影响关系(七人因素依次以1到7表示)

表3: 准则层与方案层的影响关系

生活区	工业区	山区	主干道区	公园绿地
1,3,4,6	2,3,4,6	3,5,6	3,4	3,4,5,7

### 2) 构造成对比较矩阵

比较第 $i$ 个元素与第 $j$ 个元素相对上一步某个因素的重要性时,使用数量化的相对权重 $a_{ij}$ 来描述。设共有 $n$ 个元素参与比较,则 $A = (a_{ij})_{n \times n}$ 称为成对比较矩阵。对于多指标、量化两两间的相对重要程度,取值为1/9, 1/2, 1/9, 1为同等重要,9表示 $i$ 对 $j$ 极端重要,1/9表示 $i$ 对 $j$ 极端不重要。求出该矩阵的最大特征根 $\lambda_{\max}$ 以及其对应的特征向量,将特征向量的分量归一化后,各分量就是各项指标的权重。

表4给出由第二步得到的各区的综合污染指数,转化为各区对污染的贡献的成对比较矩阵。见表5。根据以上分析得到反映各因素对各区域污染程度贡献的成对比较矩阵,见表6。

表4 各区的综合污染指数

生活区	4.993
工业区	7.737
山区	4.730
主干道地区	6.897
公园绿地区	4.866

表5 各区对污染的贡献的成时比较

	生活区	工业区	山区	主干道地区	公园绿地区
生活区	1.000	0.250	2.000	0.333	1.000
工业区	4.000	1.000	9.000	2.000	4.000
山区	0.500	0.111	1.000	0.200	0.500
主干道地区	3.000	0.500	5.000	1.000	3.000
公园绿地区	1.000	0.250	2.000	0.333	1.000

表6 各因素对各区城市污染程度贡献的成时比较

	生活区	工业区	山区	主干道地区	公园绿地区
生活区	1.000	0.250	2.000	0.333	1.000
工业区	4.000	1.000	9.000	2.000	4.000
山区	0.500	0.111	1.000	0.200	0.500
主干道地区	3.000	0.500	5.000	1.000	3.000
公园绿地区	1.000	0.250	2.000	0.333	1.000
特征根	CI	CR			
5.017	0.004	0.004			
归一权重	0.104	0.462	0.052	0.278	0.104

### 3) 计算权向量并做一致性检验

对于每一个成对比较矩阵, 计算最大特征根及对应的特征向量。定义一致性指标 CI。平均随机一致性指标 RI, 进而相除得到随机一致性比率 CR, 将其作为检验成对比较矩阵一致性的标准。若矩阵通过检验, 特征向量归一化后即为权向量。若不通过, 需重新构造成对比较矩阵。检验成对比较矩阵 A 一致性的步骤如下:

计算衡量一个成对比较矩阵  $A$  ( $n > 1$  阶方阵) 致程度的指标  $CI = \frac{\lambda_{\max} A - n}{n - 1}$ ,  $CI$  非负, 且  $CI$  越小, 一致性越好。

RI 称为平均随机一致性指标, 它只与矩阵阶数  $n$  有关, 是固定的经验值, 见表 7。

计算成对比较矩阵 A 的随机一致性比率  $CR = \frac{CI}{RI}$

判断矩阵一致性方法如下:当 $CR < 0.1$ 时,可判定成对比较矩阵 $A$ 具有较好一致性,或其不一致程度在容许范围内,否则必须调整成对比较矩阵 $A$ ,直到取得较好一致性为止。表5中各因素权重以及一致性检验情况如表6,从表中可以得知,该矩阵一致性较好。

计算组合权向量并做组合一致性检验。计算最下层各因素对目标层的权重,同理做组合一致性检验。表8是5区污染贡献权重,表9为各种污染因素的最终权重。

表7: 平均随机一致性指标与矩阵阶数 $n$ 的关系

矩阵阶数	3	4	5	6	7	8	9
RI	0.51	0.89	1.12	1.25	1.35	1.42	1.46

表8: 5区污染贡献权重

	目标层			
0.04	0.462	0.052	0.278	0.104
生活区	工业区	山区	主干道路区	公园绿地
1346	2346	356	34	3457

表9: 各种污染因素的最终权重

生活垃圾	工业生产	自然沉积	交通污染	植物吸附	人口密度	化肥农药
1	2	3	4	5	6	7
0.062	0.340	0.22	0.293	0.037	0.085	0.068

总排序一致性检验公式为

$$CR = \frac{\sum_{j=1}^4 a_j CI_j}{\sum_{j=1}^4 a_j RI_j}$$

即对第一层中采用权重矩阵计算权值的四个区域(生活区、工业区、山区、公园绿地)的 $CI/RI$ 进行加权计算。表8中若干个因素分别对四个 $\times$ 域影响, $CI$ 来自于反映这些影响的四个矩阵。 $CR$ 结果为 $0.00922 < 0.1$ ,故总层次排序一致性较好。

#### 4) 结论

最主要的影响因素是,工业生产和交通污染,其他5种污染因素严重性由大到小排序为自然沉积、人口密度、化肥农药、生活垃圾、植物吸附。



## 4 问题3模型建立与求解(略)

## 5 问题4模型建立与求解(略)

## 参考文献

- 1 百度百科. 克氏金插值 [M]. <http://baike.baidu.com/view/1684023.htm>, 2011.
- 2 百度百科. Kriging interpolation [M]. <http://baike.baidu.com/view/1684023.html>, 2011.
- 3 李柱水, 翁 月 晶. 环境地质评价原理 [M]. 北京: 化学工业出版社, 2004.  
L Z Y, Ding J, Peng J H. The Principles and Method of Environmental Quality Evaluation [M]. Beijing: Chemical Industry Press, 2004.
- 4 陈多宝, 李永梅, 柯 杰. 数学建模 [M]. 北京: 科学出版社, 2007.  
Chen D B, Li D M, Wang S Z. Mathematical Models [M]. Beijing: Science Press, 2007.
- 5 王 贵田. 重金属污染评价 [M]. 北京: 气象出版社, 2011.  
Wang J. Research on Heavy Metals in soil of Guizhou [M]. Beijing: Meteorology Press, 2011.

## Analysis of Heavy Metal Pollution in Urban Topsoil

AN Chai-kai<sup>1</sup>, HUANG Xiang-yi<sup>2</sup>, LIAN Chen<sup>2</sup>  
Advisor: DONG Zi-jing<sup>2</sup>

(1- School of Electronic Engineering and Computer Science, Peking University, Beijing 100871;  
2- School of Mathematical Sciences, Peking University, Beijing 100871.

**Abstract** During the rapid pace of the Chinese industrialization in recent years, the environmental side effect especially the soil pollution becomes more and more serious. Accumulating many data of the soil pollution, researchers pay increasing attention on unveiling the relationship between human activities and the environment by using statistical and machine learning tools. This paper evaluates the pollution distribution and analyzes the pollution level of eight major heavy metals in a certain urban area by utilizing Kriging interpolation and principal component analysis in part 1. In the hierarchical analytic way in part 2 the paper finds out the major reasons factors underlying soil pollution and substantiates the relative weights of these factors.

**Keywords** Kriging interpolation; principal component analysis; analytic hierarchy process

文章编号 1005-3085 (2011)07-0075-08

## 土壤表层重金属污染研究

王九亚, 李达超, 朱怀桦

(中山大学数学与计算科学学院, 广州 510275)

**摘 要** 本文研究在野外非均匀金属污染源分布下, 金属在土壤中的扩散. 结合重金属污染源位置与值处地中金属浓度分布的梯度变化, 利用模型, 开展数值仿真. 这是一个很有意义的问题, 值得研究者对此进行进一步的深入研究.

**关键词** 非均匀金属污染源分布; 金属在土壤中的扩散; 数值仿真. 我们通过对野外数据与数值仿真, 得到土壤金属浓度分布传播模型, 分析金属在土壤中的扩散. 在野外数据分布与数值仿真, 得到土壤金属浓度分布传播模型. 通过分析野外数据与数值仿真的结果, 我们得到金属在土壤中的扩散. 通过分析野外数据与数值仿真的结果, 我们得到金属在土壤中的扩散.

**关键词** 重金属污染; 梯度场模型; 扩散模型; 恒定污染源; 梯度污染源

**分类号** AMS: 20XX 90XX; **中图分类号** 1729; **文献标识码** A

## 1 基本假设

本模型只讨论土壤表层金属污染, 不考虑污染浓度随时间、空间的变化和扩散情况. 假设重金属只在土壤表面传播.

2. 重金属污染源代表了该地土壤中各重金属元素浓度的本底值, 实际测量值则在本底的基础上反映人为活动造成的污染程度.

3. 金属污染源总是重金属浓度的极大值点, 个别污染最严重的点可作为主要污染源.

## 2 建立模型

### 2.1 模型结构

模型结构如图1所示.

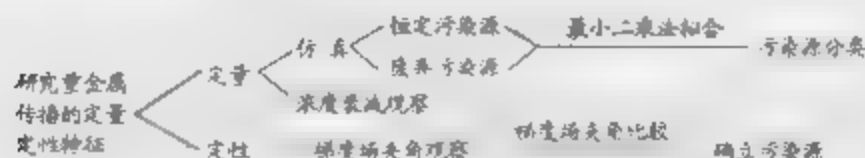


图1 分析步骤

## 2.2 模型的确立

### 2.2.1 重金属污染分布

#### 1) 重金属浓度分布

城上的地表是不规则曲面, 重金属在表层迁移, 考虑了测量方位和位置上的实际高度, 不应直接在平面上对浓度进行插值

a) 利用已知测量数据  $P(x, y)$ , 在平面网格上对海拔值进行插值, 得到城上曲面地形图

b) 根据  $z$  值, 查得此网格曲面和重金属浓度测量值  $C$ , 进行曲面的浓度插值, 得到重金属污染在曲面上的分布函数

第一步采用了 Matlab 软件样条插值方法, 此方法计算得到的曲面较为光滑, 效果较好. 第二步采用了 Kriging 插值方法进行了浓度插值<sup>[1]</sup>

#### 2) 污染级别评价

采用地累积指数法对重金属污染程度进行评级. 地累积指数法是通过计算已知浓度值  $C$  与地统计值  $Bas$  之比关系来评价人类活动造成的污染严重程度. 其计算式如下

$$I_{geo} = \frac{\log_2 C}{k \times Bas}, \quad (1)$$

其中  $k$  为修正系数, 各地  $k$  值不同, 一般令  $k$  取背景值的变动而取的系数. 一般取值为 1.5.  $I_{geo}$  为地累积指数. 根据  $I_{geo}$  将污染等级分为 6 级<sup>[2]</sup>, 如表 1 所示.

表 1: 地累积指数与污染程度分级

$I_{geo}$	< 0	0.1	1.2	2.3	3.4	4.5	> 5
级数	0	1	2	3	4	5	6
污染程度	无	无中	中	中强	强	强极强	极强

### 2.2.2 重金属污染传播特征分析

#### 1) 衰减分布函数的特征(定量)

首先, 确定一个存在污染较为严重的污染源, 即主要污染源. 设为原点  $O$ , 这样就可以排除地形和衰减的影响. 以  $O$  为中心, 划定污染级别高于 1 的  $x$  域  $D$ . 考察在  $D$  内的污染衰减分布  $d = d(r)$ , 式中  $r$  为  $D$  上任意一点与  $O$  的距离. 寻找相似分布, 以便利用这些已知的分布函数去拟合  $d(r)$ .

#### 2) 扩散模型的仿真(验证定量特征和 $k$ )

土壤上的重金属传播类似于固体与固体间的扩散模型, 故尝试使用扩散方程来描述这一过程并进行仿真试验. 考察污染源从空间上分布的衰减随时间变化的过程.

Fick's 扩散定律<sup>[3]</sup> 扩散物质在单位时间内沿该方向流过单位面积的质量与物质浓度沿该方向的导数成正比

$$\frac{\partial C(x, y)}{\partial t} = D(x, y) \left( \frac{\partial^2 C(x, y)}{\partial x^2} + \frac{\partial^2 C(x, y)}{\partial y^2} \right) \quad (2)$$

其中  $D(x, y)$  为扩散系数,  $C(x, y)$  为  $t$  时刻在  $(x, y)$  点的污染物浓度. 假设污染发生前, 该地区的重金属浓度背景值为  $b$ . 为简化问题, 以对称点扩散中心  $C(0, x_0, y_0) = C_0$  的扩散系统进行时间和空间离散化的仿真

$$C(0, x, y) = \begin{cases} C_0, & x = x_0 \wedge y = y_0, \\ b, & \text{其它.} \end{cases} \quad (3)$$





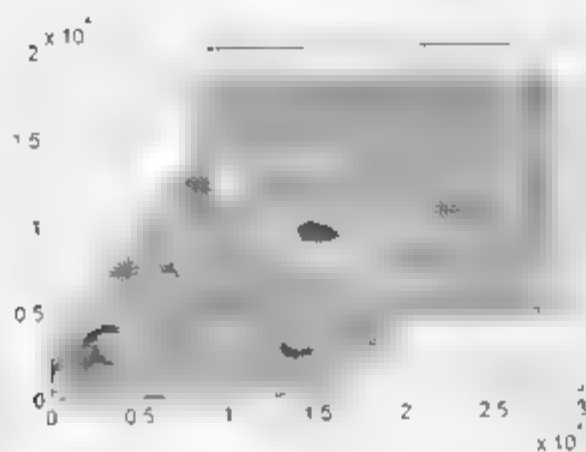


图2: 地表污染指数分布图

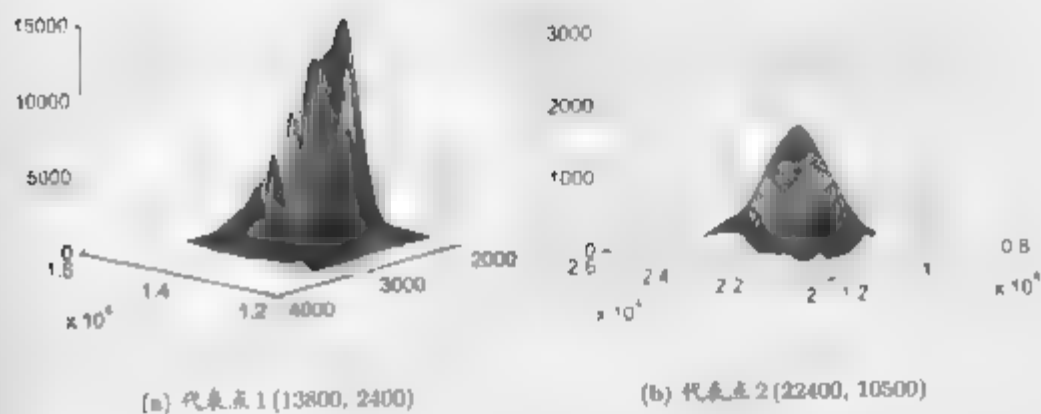


图3: 污染深度分布情况

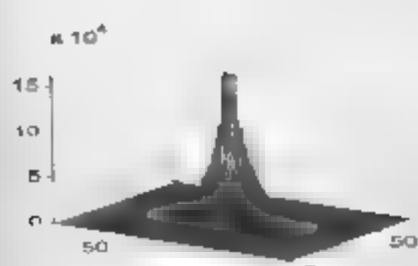


图4: 恒定污染源

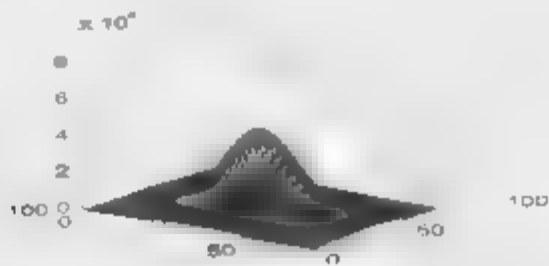


图5: 移动污染源



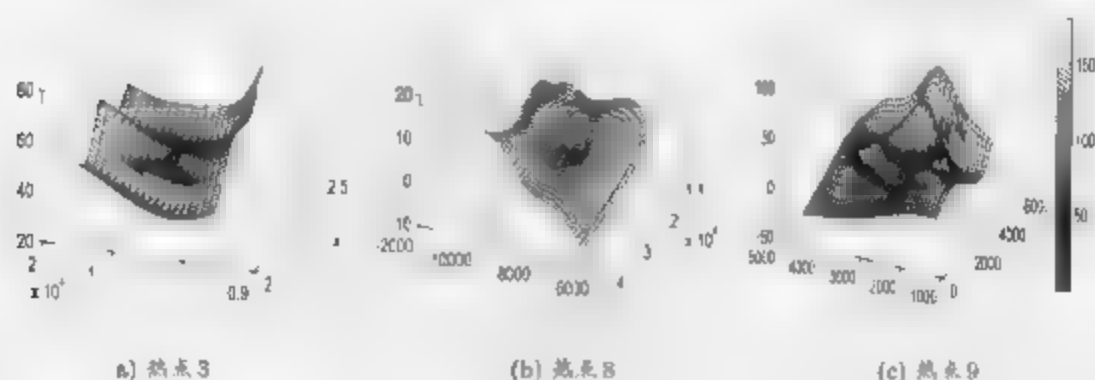


图 7: 热点处地形图

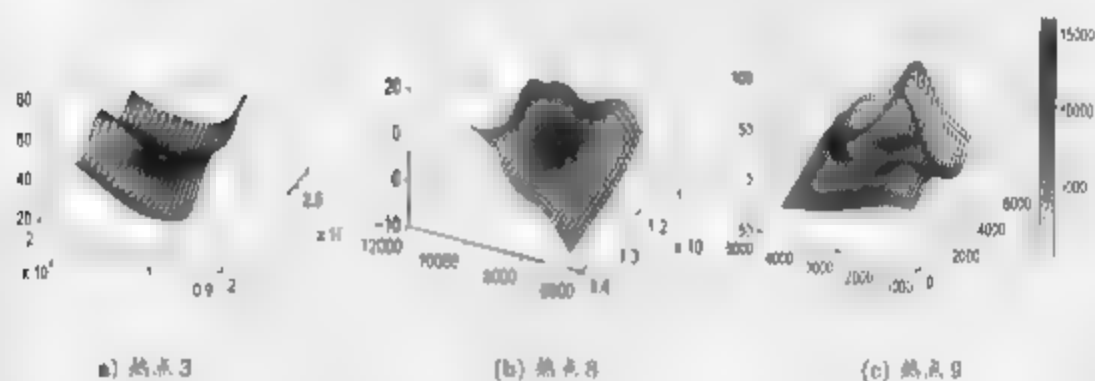


图 8: 源区地形图

### 3.2.3 污染源分类

本研究以 Hg 与 Pb 污染分布为例,使用一般点、线拟合以及端点角比较法判定污染源的类型,如表 2 所示。

表 2: 污染源分类

热点	1	2	3	4	5	6	7	8	9
$FFR_e \cdot 10^3$		2.315	0.19	0.208	0.318	0.098	2.413	3.389	2.980
$FFR_d \cdot 10^3$		2.285	0.208	0.133	0.373	0.114	3.104	3.928	3.477
污染源类型	非污染源	废弃	恒定	废弃	恒定	恒定	恒定	恒定	恒定
污染源形状	非污染源	点状	点状	点状	线状	点状	线状	线状	线状



## 4 模型评价

本文最大的优点在于开创性地提出了基于梯度场比较的方法筛选污染源,通过拟合确定污染源类型的方法。同时,我们也大胆地使用扩散模型进行仿真,得到的结果是让人意外的。不同的初始条件会造成完全不同的两种典型分布。这启发我们将污染源分为恒定污染源和废弃污染源两类。经过对 Hg 金属的分析和计算,证实这种方法非常有效。它充分发掘了地形梯度和重金属浓度梯度之间的负相关关系,更重要的是,与一般的曲面拟合方法相比,它确定时不受插值点的影响。缺点是我们还不能完全自动化运行,此过程需要借助观察图形来判断。

数据上还有大量工作有待挖掘。我们在考虑重金属传播时仅仅留在局部,没有顾及大范围重金属迁移与地形的关系。如一些污染程度较弱的通道与城市布局之间的关系等,这也是另一个很有意思的问题。

## 参考文献:

- 1 祁国英, 李. GIS在环境中的应用——以重金属污染为例[J]. 合肥: 合肥工业大学, 2006.  
Q Y H. The study on the spatial distribution and pollution assessment of heavy metal elements in soil of Tongling area based on GIS. Hefei: Hefei University of Technology, 2006.
- 2 林伯安. 地理学 GIS 中重金属污染扩散模型[J]. 长沙: 中南大学, 2009.  
Lin B A. Pollution assessment and prediction of heavy metals in soil based on Geostatistics and GIS. Changsha: Central South University, 2009.
- 3 朱 钧. 物理化学. 第三版[M]. 北京: 冶金工业出版社, 2008.  
Zhu J L. Physical Chemistry. 3rd Edition [M]. Beijing: Metallurgical Industry Press, 2008.

## Research on Heavy Metal Pollution of Topsoil

WANG Jiu-ya, LI Da-chao, ZHU Hua-jun

School of Mathematics & Computational Science, Sun Yat-sen University, Guangzhou 510275

**Abstract** An effective method for locating the pollutant source and a categorization of pollutant sources are proposed. We first obtain spread pattern of heavy metals in topsoil through analysis of real data and simulation results. With different initial values, we discover two different types of pollutant sources and the corresponding spread patterns. One is constant pollutant source and another is abandoned pollutant source. Then we locate real pollutant source by comparing the angle between gradients of floor surface and density surface. Finally we categorize these pollutant sources by using surface fitting with Gaussian distribution and exponential distribution based on the least square technique.

**Keywords:** heavy metal pollution; angle between gradients; diffusion model; constant pollutant resource; abandoned pollutant resource

Article ID:1005-3085(2011)07-0083-08

# An Approach to Predict the Location of the Contaminant Sources

LI Ze-zhe, XU Jing-wei, XU Wen-jia

(Department of Mathematics, National University of Singapore, Singapore 119076)

**Abstract.** This paper presents a mathematical model to predict the location of the contaminant sources based on the profile of the pollution level of the against region. To do the prediction, we assume certain sampling sites on the sources and simulate the pollution level of the region. Numerical mathematical tools are used to simulate the diffusion of heavy metal through sea and air. By varying the assumption of the sources and comparing the simulated situation under different assumptions with the real situation, we obtain the optimal number and location of the sources. Computer implementation of the model gives a reasonable result.

**Keywords:** contaminant sources location; heavy metal pollution

**Classification:** AMS(2000) 65K05    **CLC number:** O29    **Document code:** A

## 1 Introduction

Nowadays, heavy metal pollution becomes increasingly prevalent in our daily life. In some areas, it has a fairly negative impact on our health. To reduce the pollution, we need to accurately locate the contaminant sources. In this paper, we offer a pragmatic way to analyse the pollution data and predict the location of contaminant sources.

### 1.1 Outline of our paper

- 1) Section 2: concept of pollution index;
- 2) Section 3: theoretical fundament of our model;
- 3) Section 4 and 5: results, conclusion, assessment and generalisation of our method.

### 1.2 General assumptions

- 1) The density of heavy metal of the sample site is the average density of the square kilometre.
- 2) The land shape can be approximated by the smooth surface connecting all the sample sites.
- 3) Only diffusion through soil and diffusion through air are considered in our model. Other types of propagation may exist, but are not statistically significant.
- 4) The source of contaminant is located at one of the sample sites.

## 2 Pollution index

To measure the level of heavy metal pollution for each site, we adopt the Nemerow pollution index<sup>[1]</sup>

Single contamination index  $P_i = (C_i/S_i)$ , where  $P$  is the single contamination index of heavy metal pollutant  $i$ ,  $C_i$  is its real concentration, and  $S_i$  is its regional background value.

Nemerow pollution index<sup>[2]</sup>

$$P_c = [(P_{\max}^2 + P_{\text{avg}}^2)/2]^{\frac{1}{2}},$$

where  $P_c$  is the Nemerow pollution index of the site,  $P_{\max}$  is the highest single contamination index among different heavy metals, and  $P_{\text{avg}}$  is the average single contamination index.

### 3 Mathematical model to predict the location of contaminants

In this section, we present the theoretical work of our model.

We will first assume one of the sites in the region to be the contaminant source and simulate the pollution level of the entire region under this assumption. Normal distribution is used to approximate the distribution of heavy metals in a flat region. Several correction factors, such as the altitude difference, the smoothness of land path, the direction and strength of wind, are taken into consideration. Then, we compare the simulated pollution level of the region with the real situation provided in the excel file to test the degree of similarity (using mean square error). We then change our assumption about the contaminant sources and do the simulation. After all possible assumptions are simulated, the assumption with the greatest degree of similarity is concluded to be the location of the contaminant sources.

#### 3.1 Sub-model one: diffusion through soil

As the real world operates, normal distribution is the most general objective distribution law. Thus, it is reasonable to use normal density function to approximate the distribution of heavy metal in a flat region.

$$f_A(B) = \frac{1}{\sqrt{2\pi}\sigma_B} \exp\left(-\frac{d_{A,B}^2}{2\sigma_B^2}\right),$$

where  $A$  is the assumed contaminant source,  $B$  is the point whose level of pollution is to be simulated under the assumption that  $A$  is the contaminant source,  $d_{A,B}$  is the distance between  $A$  and  $B$ , and  $f_A(B)$  denotes the pollution concentration at  $B$ .

Since the region under investigation is far from a flat region, we need to take into account the relief of the terrain between  $A$  and  $B$ . Thus, we introduce two modifications to the above formula: smoothness and difference in altitude.

Since the rougher the land path between  $A$  and  $B$ , the more difficult for heavy metals to propagate, we use the correlation coefficient  $r$  to measure the smoothness of the land path and modify the formula as follows:

$$F_A(B) = r \times f_A(d_{A,B})$$

The correlation coefficient is calculated by using the point  $B$  (the contaminant source point  $A$  and all other site points whose  $x$ - $y$  coordinates lie near ( $<1$  km) the line connecting  $A$  and  $B$ ):

$$r = \frac{\sum_{i=1}^n (h_i - \bar{h})(d_i - \bar{d})}{\sqrt{\sum_{i=1}^n (h_i - \bar{h})^2 \sum_{i=1}^n (d_i - \bar{d})^2}}$$

Here we assume there're  $n$  site points lying near the line  $AB$ . Then we get the projection of these points on the vertical plane containing  $A$  and  $B$ , and denote them as  $x_1, x_2, \dots, x_n$ . For  $x_i (i = 1, 2, \dots, n)$ ,  $h_i$  is its altitude, and  $d_i$  is the horizontal distance between this point and  $A$ .

Besides the above modification, we also need to consider the difference in altitudes of  $A$  and  $B$ . Here we need to use slope  $k$  and the oblique angle  $\theta$ :

$$k = \frac{h_B - h_A}{d_{AB}}, \quad \theta = \arctan(k)$$

Since we just compare the final difference between the simulated pollution and real pollution to predict the location of the contaminant source, we can just use an assumed function that has the same properties of the real distribution function to calculate.

As far as it is concerned, when the terrain between  $A$  and  $B$  is a straight line and  $A$  is above  $B$  ( $\theta < 0$ ), the steeper the slope is, the wider the pollutants are able to spread. Accordingly, the larger  $\theta$  is, the greater  $\sigma_B$  will be. Besides, when ( $\theta < 0$ ),  $\sigma_B > \sigma$ . The analysis of situation when  $A$  is lower than  $B$  ( $\theta > 0$ ) is similar. What's more, when ( $\theta = 0$ ),  $\sigma_B = \sigma$  according to the definition, while when  $\theta$  goes to infinity, the pollution will not spread upward, and thus  $\sigma_B = 0$ . Then we can modify  $\sigma_B$  by  $\theta$ :

$$\sigma_B = \sigma \cos \theta, \quad \theta \geq 0, \quad \sigma_B = \frac{\sigma}{\cos \theta}, \quad \theta < 0$$

Here  $\sigma$  is the standard deviation of the normal distribution where the land is plain.

The  $\sigma_B$  function we define obeys all those properties mentioned above. So it is reasonable and well-defined.

Therefore, we obtain the final distribution function of the heavy metal pollution propagated via soil:

$$F_A(B) = |r| \times \frac{1}{\sqrt{2\pi}\sigma_B} \exp \left\{ -\frac{d_{A,B}^2}{2\sigma_B^2} \right\}$$

### 3.2 Sub-model two: diffusion through air

In this sub-model, we still assume that the distribution of pollution level can be approximated by a normal distribution centered at the contaminant source. However, we need to take into consideration some other influencing factors: wind strength and wind direction, which may change the shape of the distribution.

First, assume that the direction of wind is completely random, and then the distributions of the concentrations of heavy metals will follow normal distribution, which centres at the original point (contaminant source), and thus  $x \sim N(0, \delta^2)$ .

Then based on the analysis of the spatial distribution of pollution provided in the exercises, we find out that the distribution of heavy metals is severely skewed towards the left-bottom corner of the graph. Thus, it is reasonable for us to assume that the direction of wind follows the direction of the vector  $(-1, -1)$ . We assume that the strength of wind is random over a long period. Thus, the wind will shift the normal distribution curve. Hence the mean of the final normal distribution is approximately the uniform distribution  $\mu \sim (-L_{\text{average strength}}, 0)$ , where  $L_{\text{average strength}}$  is the distance that the wind with average strength

can flow the substances containing heavy metals forward in the direction of the wind. Thus, we can calculate the distribution along the direction  $x = 1 \sim 2$  by stochastic analysis.

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}\delta} \exp\left[-\frac{(x-\mu)^2}{2\delta^2}\right], \quad f(\mu) = \frac{1}{L_{\text{average strength}}}$$

Then

$$f(x) = \int_0^{L_{\text{average strength}}} f(x, \mu) d\mu = \frac{\sqrt{\pi} \left( \operatorname{Erf}\left[\frac{L_{\text{average strength}} - x}{\sqrt{2}\delta}\right] + \operatorname{Erf}\left[\frac{x}{\sqrt{2}\delta}\right] \right)}{2\sqrt{\pi}\delta L_{\text{average strength}}}$$

where

$$\operatorname{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

Then we need to know how it will change the mean and variance of the normal distribution near any variable, which show how the wind influences the location and spread of the pollution. We attain this aim in the following way.

$$E(X|\mu = \mu_0) = E(N(\mu = \mu_0, \delta^2) | \mu = \mu_0) = \mu$$

Hence

$$E(X) = E(E(X|\mu)) = E(N(\mu, \delta^2)) = E(\mu) = \frac{L_{\text{average strength}}}{2}$$

Then

$$\operatorname{Var}(X) = E(\operatorname{Var}(X|\mu)) + \operatorname{Var}(E(X|\mu)),$$

$$\operatorname{Var}(X, \mu = \mu_0) = \operatorname{Var}(N(\mu = \mu_0, \delta^2) | \mu = \mu_0) = \delta^2$$

Thus we have

$$\operatorname{Var}(X) = E(\operatorname{Var}(X|\mu)) + \operatorname{Var}(E(X|\mu)) = E(\delta^2) + \operatorname{Var}(\mu) = \delta^2 + \frac{L_{\text{average strength}}^2}{12}$$

The above analysis is based on two points which are exactly in the same position as the wind vector  $\vec{e} = (-1, -1)$ . Then we will generalize the distribution cases as any two points  $S$  (phase  $\alpha$ ) to a stationary point  $A(x_a, y_a)$  and other point  $B(x_b, y_b)$  (Fig. 1) there are two vectors, namely  $\vec{e} = (-1, -1)$ ;  $\vec{AB} = (x_b - x_a, y_b - y_a)$ .

We can then calculate the angle between these two vectors

$$\cos \alpha = \frac{(\vec{AB} \cdot \vec{e})}{|\vec{AB}| |\vec{e}|}$$

Knowing the angle between these two vectors, we can obtain a new random variable which suits the general cases by changing the uniform distribution  $\mu = 1 - 0 \cos \alpha (L_{\text{average strength}})$  and we plug it into the previous equations in the following way.

$$f(x) = \frac{\sqrt{\pi} \left( \operatorname{Erf}\left[\frac{L_{\text{average strength}}(1 - \cos \alpha) - x}{\sqrt{2}\delta}\right] + \operatorname{Erf}\left[\frac{x}{\sqrt{2}\delta}\right] \right)}{2L_{\text{average strength}} \cos \alpha}, \quad f(X) = \frac{L_{\text{average strength}}}{2} \cos \alpha$$

$$\operatorname{Var}(X) = E(\operatorname{Var}(X|\mu)) + \operatorname{Var}(E(X|\mu)) = E(\delta^2) + \operatorname{Var}(\mu) = \delta^2 + \frac{L_{\text{average strength}}^2}{12} \cos^2 \alpha$$

By using this method, we can calculate the concentration of the ending point by knowing the contaminant point given that it diffuses only through air.

### 3.3 Combination of the sub-models

From the above two sub-models, we obtain the distribution of the heavy metal pollution propagated via soil and the distribution of the pollution diffused via air. Now we combine these two distributions to get the final distribution of the pollution.

To calculate the final distribution, we assume a fixed proportion of the pollution propagated via soil. Accordingly, 1 -  $\rho$  of the pollution is diffused by air.

Assuming A is the source, we want to get the concentration of the pollution at B. Firstly, we need to obtain the ratio of the pollution concentration at B to the concentration at A diffused via soil and air respectively by the two sub-model. We use  $C_A$ ,  $C_A'$ ,  $C_B$ ,  $C_B'$  to denote the pollution concentration diffused via soil and air at A and B respectively.

According to the sub-models, we have

$$C_B = \frac{C_A \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}}} = C_A \times \frac{e^{-\frac{1}{2\sigma^2}}}{e^{-\frac{1}{2\sigma^2}}} = C_A$$

$$C_B' = \frac{C_A' \times \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\frac{1}{2\sigma'^2}}}{\frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\frac{1}{2\sigma'^2}}} = C_A' \times \frac{e^{-\frac{1}{2\sigma'^2}}}{e^{-\frac{1}{2\sigma'^2}}} = C_A'$$

$$C_B = \frac{C_A \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}}} = C_A \times \frac{e^{-\frac{1}{2\sigma^2}}}{e^{-\frac{1}{2\sigma^2}}} = C_A$$

$$C_B' = \frac{C_A' \times \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\frac{1}{2\sigma'^2}}}{\frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\frac{1}{2\sigma'^2}}} = C_A' \times \frac{e^{-\frac{1}{2\sigma'^2}}}{e^{-\frac{1}{2\sigma'^2}}} = C_A'$$

The concentration of the total pollution concentration at B given A is the pollution concentration

$$\rho = \frac{C_B}{C_A} = \frac{C_B}{C_A} \times 1 = \frac{C_B}{C_A}$$

where the estimated pollution concentration at B is  $C_B = \rho \times C_A$ , where  $C_A$  is the real pollution concentration we collect at A.

According to 1, we use Nemerow pollution index to represent the pollution. Hence, the formula where  $P_B^A = \rho \times P_A$ , where  $P_B^A$  is the estimated Nemerow pollution index at B given A is the contaminant source, and  $P_A$  is the real Nemerow pollution index at A.

By improving the above method, we can get the dataset of the estimated Nemerow pollution indexes of pollution for the assumption that a certain point is the contaminant source. For instance, we assume point i is contaminant source, and then we get  $P_k^i$ , which denotes the estimated Nemerow pollution index of pollution at point k given i is the source. So we can get 319 such  $P_k^i$ s.

To compare the estimated value with the real one, we use the following formula

$$MSE_k = \frac{\sum_{k=1}^{319} (P_k^i - P_k)^2}{319}$$

where  $P_k$  is the real pollution concentration at point k.

Applying the same method to every point (changing i we mentioned in the last paragraph), we can get 319 datasets of the estimated values, and we also get 319 MSEs.

If there is only one contaminant source, we shall find out  $\min_i \{MSE_i\}$  when  $i = 1$ , and then point  $i$  is the contaminant source.

If there are two contaminant sources, we use  $P_k^i$  ( $i = 1, 2, \dots, 319$ ) generated above to calculate the sum pollution concentration caused by two contaminant sources. For example, we assume points 1, 2 are two contaminant sources, and then we get the estimated pollution concentration at point  $k$

$$P_k^{1,2} = P_k^1 + P_k^2$$

Similarly, we get

$$MSE_{i,j} = \frac{\sum_{k=1}^{319} (P_k^{i,j} - P_k)^2}{319}$$

and find out  $\min_{i,j} \{MSE_{i,j}\}$  when  $i = m, j = n$ , and points  $m$  and  $n$  are the most possible two contaminant sources.

By using the same method we could get the most possible  $k$  contaminant sources ( $k = 1, 2, 3, \dots$ ). But we cannot increase  $k$  to 319, which will be too large for programming. So we just increase  $k$  until the minimum MSE of  $k = 1$  assumed sources is smaller than the minimum MSE of  $k$  assumed sources. Thus, we get the most possible number of the contaminant sources and the location of them.

By combining the two sub-models, we can predict the location of the contaminant sources.

## 4 Results & interpreting the results

### 4.1 Results

According to 'Evaluation Soil Contamination' published by U.S. department of interior<sup>[3]</sup>, it is reasonable to assume that  $\tau_{As} = 0.38$ ,  $\tau_{Cd} = 0.18$ ,  $\tau_{Cr} = 0.64$ ,  $\tau_{Cu} = 0.42$ ,  $\tau_{Hg} = 0.21$ ,  $\tau_{Ni} = 0.17$ ,  $\tau_{Pb} = 0.08$ ,  $\tau_{Zn} = 0.35$  and we assume that  $\sigma = 500$ ,  $\delta = 5000$ . The results of the estimated contaminant sources and the related minimum MSEs are listed in Tables 1, 2, 3.

Table 1: Four possible contaminant sources

Four contaminant sources				Minimum MSE
(3299, 6018)	(2383, 3617)	(13694, 2357)	(15248, 9106)	1892.324
(2383, 3617)	(15248, 9106)	(13694, 2357)	(13797, 9621)	2498.879
(15248, 9106)	(13694, 2357)	(9319, 6799)	(2383, 3617)	2754.345
(13694, 2357)	(13797, 9621)	(2383, 3617)	(3299, 6018)	3564.231

Table 2: Five possible contaminant sources

Five contaminant sources					Minimum MSE
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(3299, 6018)	1094.358
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(13797, 9621)	1732.398
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(9319, 6799)	1921.301
(2383, 3617)	(15248, 9106)	(13694, 2357)	(3299, 6018)	(13797, 9621)	2487.628

Table 3: Six possible contaminant sources

Six contaminant sources	Minimum MSE
(2383, 36.7), (5248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (13797, 9621)	1108.231
(2383, 36.7), (5248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (9319, 6799)	1324.387
(2383, 36.7), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (8629, 12086)	1431.216
(2383, 36.7), (15248, 9106), (13694, 2357), (2708, 2295), (13797, 9621), (9319, 6799)	442.418

## 4.2 Interpreting the result

From the mean square error, it is seen that when there are five contaminant sources, we achieve the minimum mean square errors and their locations are (2383, 3617), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018). This is reasonable because all of these five points are frequently shown in the above tables, indicating that they are of a large probability to be the contaminant points.

However, since the mean square error is still very large, we have to suspect our result. Because we have made an assumption which states that the only possible contaminant points are those points shown in the data file, but in reality, it may not be correct. If we revise it a bit and try the middle points between (2383, 3617) and (2708, 2295) and also substitute the middle point of their heavy metals concentrations, we can obtain a better mean square error which is 190.134, 5 times smaller than the previous mean square error when we try 3 contaminant points. These three contaminant points are (15248, 9106), (13694, 2357), (2545.5, 2956).

## 5 Additional information & evolution model

If time factor can be taken into consideration, we could complete our work to develop an evolution model for the geological environment of the city. Hence we need to collect the concentrations of heavy metals at different times at the same points as we choose previously. Analyzing the dynamic data, we may obtain the accurate  $\sigma$ . As time goes by, we are also able to collect the weather data (weather of the city) so as to analyze the impact on pollution distribution by rainfalls, the wind direction and the information about the rivers across the city.

Given the information mentioned above, we could modify our model by correcting the pollution extent and wind direction in diffusion by air sub-model, adding the factor of the rivers and rainfalls. Furthermore, we can develop a new model to study the evolution models of geological environment of the city on the basis of the current model. In the new model, we could analyze the pollution distribution of a certain point as time goes on, instead of the pollution distribution in a large area at a certain time. This new model will also help us predict the trend of the pollution distribution as time goes on. What's more, we could predict what the environment we reside will be in the future, and the significance of the environmental protection will become a highlight of our life. Therefore, we are able to avoid the trend of the environment deterioration, and make our environment better and better.



If there is only one contaminant source we shall find out  $\min_k \{MSE_k\}$  when  $k = 1$ , and then point 1 is the contaminant source.

If there are two contaminant sources, we use  $P_k^i$  ( $i = k = 1, 2, \dots, 319$ ) generated above to calculate the sum pollution concentration caused by two contaminant sources. For example, we assume points 1, 2 are two contaminant sources, and then we get the estimated pollution concentration at point  $k$ :

$$P_k^{1,2} = P_k^1 + P_k^2$$

Similarly, we get

$$MSE_{i,j} = \frac{\sum_{k=1}^{319} (P_k^{i,j} - P_k)^2}{319},$$

and find out  $\min_{i,j} \{MSE_{i,j}\}$  when  $i = m, j = n$ , and points  $m$  and  $n$  are the most possible two contaminant sources.

By using the same method we could get the most possible  $k$  contaminant sources  $k = 1, 2, 3, \dots$ . But we cannot increase  $k$  to 319 which will be too large for programming. So we just increase  $k$  until the minimum MSE of  $k-1$  assumed sources is smaller than the minimum MSE of  $k$  assumed sources. Thus, we get the most possible number of the contaminant sources and the location of them.

By combining the two sub-models, we can predict the location of the contaminant sources.

## 4 Results & interpreting the results

### 4.1 Results

According to Evaluation Soil Contamination published by U.S. department of interior<sup>[3]</sup>, it is reasonable to assume that  $\tau_{As} = 0.38$ ,  $\tau_{Cd} = 0.18$ ,  $\tau_{Cr} = 0.64$ ,  $\tau_{Cu} = 0.42$ ,  $\tau_{Hg} = 0.21$ ,  $\tau_{Ni} = 0.17$ ,  $\tau_{Pb} = 0.08$ ,  $\tau_{Zn} = 0.35$ , and we assume that  $\sigma = 5000$ ,  $\delta = 5000$ . The results of the estimated contaminant sources and the related minimum MSEs are listed in Tables 1-2-3.

Table 1- Four possible contaminant sources

Four contaminant sources				Minimum MSE
(3299, 6018)	(2383, 3617)	(13694, 2357)	(15248, 9106)	1892.324
(2383, 3617)	(15248, 9106)	(13694, 2357)	(13797, 9621)	2498.879
(15248, 9106)	(13694, 2357)	(9319, 6799)	(2383, 3617)	2754.345
(13694, 2357)	(13797, 9621)	(2383, 3617)	(3299, 6018)	3564.231

Table 2- Five possible contaminant sources

Five contaminant sources					Minimum MSE
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(3299, 6018)	1094.358
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(13797, 9621)	1732.398
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(9319, 6799)	1921.301
(2383, 3617)	(15248, 9106)	(13694, 2357)	(3299, 6018)	(13797, 9621)	2487.628

Table 3: Six possible contaminant sources

Six contaminant sources	Minimum MSE
(2383, 36.7), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (13797, 9621),	1108.231
(2383, 36.7), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (9319, 6799)	1324.387
2383, 3617, (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018), (8629, 12086)	1431.216
(2383, 3617), (15248, 9106), (13694, 2357), (2708, 2295), (13797, 9621), (9319, 6799)	1442.418

## 4.2 Interpreting the result

From the mean square error, it is seen that when there are five contaminant sources, we achieve the minimum mean square errors and their locations are (2383, 3617), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018). This is reasonable because all of these five points are frequently shown in the above tables, indicating that they are of a large probability to be the contaminant points.

However, since the mean square error is still very large, we have to suspect our result. Because we have made an assumption which states that the only possible contaminant points are those points shown in the data file, but in reality, it may not be correct. If we revise it a bit and try the middle points between (2383, 3617) and (2708, 2295) and also substitute the middle point of their heavy metals concentrations, we can obtain a better mean square error which is 90.134, 5 times smaller than the previous mean square error when we try 3 contaminant points. These three contaminant points are (15248, 9106), (13694, 2357), (2547.5, 2856,

## 5 Additional information & evolution model

If time factor can be taken into consideration, we could complete our work to develop an evolution model for the geological environment of the city. Hence we need to collect the concentrations of heavy metals at different times at the same points as we choose previously. Analyzing the dynamic data, we may obtain the accurate  $\sigma$ . As time goes by, we are also able to collect the weather data (weather of the city), so as to analyze the impact on pollution distribution by rainfalls, the wind direction and the information about the rivers across the city.

Given the information mentioned above, we could modify our model by correcting the pollution extent and wind direction in diffusion by air sub-model, adding the factor of the rivers and rainfalls. Furthermore, we can develop a new model to study the evolution models of geological environment of the city on the basis of the current model. In the new model, we could analyze the pollution distribution of a certain point as time goes on, instead of the pollution distribution in a large area at a certain time. This new model will also help us predict the trend of the pollution distribution as time goes on. What's more, we could predict what the environment we reside will be in the future, and the significance of the environmental protection will become a highlight of our life. Therefore, we are able to avoid the trend of the environment deterioration, and make our environment better and better.

## References:

- [1] Beyer W N. Evaluation soil contamination[S]. Department of Interior. Fish and Worldwide Service. Washington. 1990.
- [2] Jiang Y et al. Contamination, source identification, and risk assessment of polycyclic aromatic hydrocarbons in agricultural soil of Shanghai, China[C]. Environmental Monitoring and Assessment, 2011, 183: 149-160.
- [3] Lian F, Wang Y X. Single-factor and Nemerow multi-factor index to assess heavy metals contamination in soils on railway side of Harbin-Suifenhe railway in Northeastern China[J]. Applied Mechanics and Materials, 2011, 71-78: 3033-3038.

## 重金属污染源估测模型

李泽彪, 许经纬, 徐文彬

(新加坡国立大学数学系, 新加坡 119076)

**摘要** 本文提供了一个根据污染数据推测污染源位置的数学模型。首先假定某个特定位置的污染数据采集点作为污染源。将污染物传播方式分为土壤传播和空气传播; 然后运用多种数学方法综合两种传播方式和污染分布进行模拟。最后通过改变污染源位置的假设, 并将每种假设下的模拟污染分布与污染数据分布进行比较, 从而得到最可能的污染源数量和位置。该过程可用计算机实现。

**关键词** 污染源; 土壤; 重金属污染



商。进而,以背景值为参照标准,将不同元素在环境中的含量,并结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

## 4 模型建立与求解

### 4.1 问题一

由表 1 中给出的数据,将 8 种元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

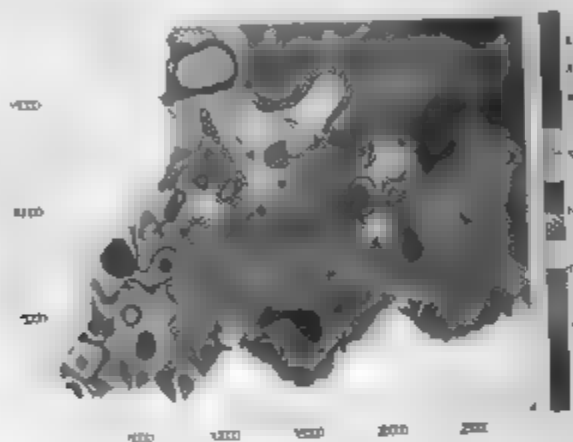


图 1: As 元素的分布图

通过表 1 中给出的数据,将 8 种元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

$$P_{ij} = \frac{X_{ij} - B_{max}}{B_{max}}, \quad j = 1, 2, \dots, 8, \quad i = 1, 2, 3, 4, 5, \quad (1)$$

计算 8 个元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

$$I_{geo} = \log_2 \frac{C_n}{1.5B_n}, \quad (2)$$

表 1 中 8 个元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

通过表 1 中给出的数据,将 8 种元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

进而得到各元素在环境中的含量,结合地壳丰度数 $Z$ ,以获得各元素的相对富集程度。依次进行,根据各元素的富集程度,通过不同元素的相对富集程度,分析不同元素在环境中的富集程度。最后,通过不同元素的富集程度,分析不同元素在环境中的富集程度。

生活区:  $Cu > Zn > Hg \geq Pb \geq Cr > Cd > As$ ;

工业区:  $Hg > Cu > Zn > Cd > Pb > As > Cr > Ni$ ;

交通区:  $Hg > Cu > Zn > Cd > Pb > Cr$ ;

公园绿地区:  $Hg > Cu > Zn > Cd > Pb > As$ ;

山区: 无重金属元素污染

综合分析,各功能区的污染程度从大到小依次为:工业区、交通区、生活区、公园绿地区、山区。

表1 各区域各重金属元素的浓度指数

区域	元素	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
生活区	$X_{ij}$	0.27	289.96	69.02	49.40	93.04	18.34	69.11	237.01
	$P_{ij}$	0.16	0.53	0.41	1.42	0.82	-0.08	0.61	1.44
	$I_{geo}$	0.22	0.57	0.57	[1.32]	0.83	-0.01	0.57	[1.20]
工业区	$X_{ij}$	7.25	393.11	53.41	127.54	642.36	19.81	93.04	277.93
	$P_{ij}$	0.34	1.07	0.09	5.23	11.60	-0.00	1.16	1.87
	$I_{geo}$	0.43	[1.01]	0.20	[2.69]	[3.61]	0.10	[1.00]	[1.43]
山区	$X_{ij}$	4.04	152.32	38.96	17.32	40.96	15.45	36.56	73.29
	$P_{ij}$	-0.25	0.20	-0.20	-0.15	-0.20	-0.22	-0.15	-0.24
	$I_{geo}$	-0.42	-0.36	-0.26	-0.19	-0.36	-0.28	-0.35	-0.50
交通区	$X_{ij}$	5.71	360.01	58.05	62.21	446.82	17.62	63.53	242.85
	$P_{ij}$	0.06	0.89	0.18	2.05	7.760	-0.11	0.48	1.50
	$I_{geo}$	0.08	[0.88]	0.32	[1.65]	[3.09]	-0.07	0.45	[1.23]
公园绿地区	$X_{ij}$	6.26	280.54	43.64	30.19	114.99	15.29	60.71	154.24
	$P_{ij}$	0.16	0.48	-0.11	0.48	1.25	-0.23	0.41	0.59
	$I_{geo}$	0.21	0.52	-0.09	0.61	[1.13]	-0.27	0.38	0.58

表2 污染程度与等级

地累积指数 $I_{geo}$	分级	污染程度	地累积指数 $I_{geo}$	分级	污染程度
$I_{geo} \geq 5$	6	极重污染	$1 \leq I_{geo} < 2$	2	中污染
$4 < I_{geo} < 5$	5	重污染 极重污染	$0 < I_{geo} < 1$	1	无污染 中污染
$3 \leq I_{geo} < 4$	4	重污染	$I_{geo} < 0$	0	无污染
$2 \leq I_{geo} < 3$	3	中污染 重污染			



殊性,其传播速度较慢。且不同重金属的传播速度不同。由此就可以得到污染源的特征。污染元素的含量与背景含量相比,相对较大,即其为区域极值。

#### 4.3.2 建立重金属污染物传播的数学模型

$$\frac{\partial u}{\partial t} + \nabla q = 0, \quad (3)$$

其中  $u$  表示重金属的浓度,  $\nabla q$  表示传播强度的损失率,由

$$q = -k \cdot u \cdot \nabla h, \quad (4)$$

可得

$$\frac{\partial u}{\partial t} = -k \nabla h \nabla u - k u \nabla^2 h = 0. \quad (5)$$

因为重金属污染物的传播不仅与时间  $t$  有关,而且与空间向量  $x$  有关。令

$$\frac{dx}{dt} = -k \nabla h, \quad (6)$$

进一步得

$$\frac{du}{dt} = k u \nabla^2 h. \quad (7)$$

进而可解得上述对流模型的解为

$$u(x, t) = U_T \exp \left( \int_0^t k \nabla^2 h(x(\tau)) d\tau \right), \quad (8)$$

式 (8) 即为重金属污染物传播的数学规律。

#### 4.3.3 确定污染源的位置

根据所有数据的特征,建立如 1 微分传递模型来寻找其污染源。由 (8) 可知,将其寻找污染源的问题转化为寻找函数空间极值的问题。

某,假设给定  $n$  个样本值,任意选取一个插值节点  $u(p, T)$ , 沿  $p$  的特征曲线反向计算  $u(x, t) / p$  ( $0 < t < T$ ), 设定一个污染浓度阈值  $u^*$ , 对于任意的  $t$  ( $0 < t < T$ ), 如果满足如下条件

$$B(t) = \{x(t) | u(x(t), t; p) \geq u^*\}, \quad (9)$$

则表明  $t$  就是可能的污染源位置。

数值实验。对于每一种重金属元素,任取一个样本点,通过上述原理,得出各种重金属污染源可能的位置,见表 5。

表 5: 各种重金属污染源的可能位置

元素	样本点 $u(p, T)$	阈值 $u^*$	污染源可能的位置
As	(15918.88, 10954.90)	15	(18234.36, 10954.09) 交通区
Cd	(11577.37, 8071.43)	500	(20549.83, 11530.62) 生活区 (26041.90, 3747.46) 工业区
Cr	(7235.85, 8071.43)	100	(2894.34, 6630.10) 生活区 (4052.08, 4035.71) 工业区
Cu	(10130.20, 8071.44)	150	(19970.96, 9224.50) 生活区 (4052.08, 3170.92) 工业区
Hg	(6656.98, 7206.84)	500	(1157.74, 3170.92) 交通区 (13892.85, 9512.77) 工业区
Ni	(6078.12, 6918.37)	20	无
Pb	(11287.93, 3747.45)	150	(14182.28, 7994.91) 生活区 (4052.08, 3170.92) 工业区
Zn	(8393.59, 6341.84)	400	(9840.76, 3747.45) 交通区 (13024.54, 10605.82) 生活区



基于样本点的数据,通过(8)式进行反演计算,得出各种重金属污染源的可能位置。由表5可以看出,各种重金属污染源主要集中在工业区和生活区,且污染源处的重金属浓度普遍较高。但是也有例外,比如Pb元素,得到其中一个可能的污染源位置为(14182.28,7994.91),属于生活区,而该点处的浓度值为38.78,低于该种元素的平均浓度浓度61.7。出现这种现象可能的原因是该种元素所在区域产生了富集现象。另外,Ni元素并没有找到污染源,这可能和整个城区Ni元素本身并没超标有关。

#### 4.4 问题四

##### 4.4.1 模型评价

本文基于所给样本数据,得到重金属的空间分布情况,并结合重金属污染物的传播特性,建立起重金属污染物的微分传播扩散模型,进而根据所给的样本点,通过反演计算,确定了可能的污染源位置,为有关重金属污染问题的解决提供了可行的决策方案。但是,由于所给样本信息有限,以及模型建立过程上的假设理想化,问题分析还存在诸多不足之处。

1) 本文基于克里格插值法,得到了各种重金属元素浓度的空间分布图。然而,由于采样数据过于离散化、数据量小等问题,导致所得空间分布可能存在很大误差,结果的准确性受到很大影响。

2) 本文基于对题设所给数据的分析,分别对各重金属元素污染的主要原因进行了定量、定性分析。但是,分析过于简单。未考虑各元素之间的关联性。

3) 本文基于重金属污染物的传播特性,通过简化假设,建立起其传播的数学模型。由于污染物传播本身的复杂性,导致污染源的位置确定存在很大误差。

##### 4.4.2 建议进一步收集的信息及改进思路

1) 增加样本数据容量。通过增加样本数据容量,从而提高插值精度,为准确判断各空间位置的金属浓度提供可靠的依据。

2) 增加其他一般元素的浓度。进行比较分析研究<sup>[5]</sup>。

3) 增加时间和序列上的样本数据。例如,监测连续五年同一空间坐标处的重金属浓度值,可考虑建立时间序列模型。通过此模型不仅可以很好的确定污染源的位置,还可以很好的研究重金属的传播以及城市地质环境的历史演变规律。

4) 增加受污染的环境变化、病力等数据。为更全面的分析提供准确可靠的依据。

#### 参考文献:

- Adison Surfer lesson[OL]. <http://wenku.baidu.com/view/8623e12a4b73f242336c5f64.html>
- 姜礼尚等.数学模型(第3版)[M].北京:高等教育出版社,2003.  
Jiang Q Y et al. Mathematical Model (3rd Edition) [M]. Beijing: Higher Education Press, 2003.
- 刘引等.地质指数法及生态危害指数法在土壤重金属污染种的应用及探讨[J].安徽农业科学,2007,35(13): 13-17.  
Peng J et al. Application of the index of geo-accumulation index and ecological risk index to assess heavy metal pollution in soils[J]. Guangdong Trans Elements Science, 2007, 14(8): 13-17.
- 姚迪等.青岛城区土壤重金属环境地球化学研究[J].中国地质,2008,35(3): 539-549.  
Yao D et al. Environmental geochemistry of heavy metals in urban soils of Qingdao city[J]. Geology in China, 2008, 35(3): 539-549.
- 王义海.土壤重金属污染与防治[J].安徽农业科学,2005,11(7): 46-47.  
Wang Y M. Soil heavy metal pollution and its prevention and cure[J]. Anhui Agricultural Science Bulletin, 2005, 11(7): 46-47.

## Research on Heavy Metal Pollution in Urban Topsoil

ZHENG Ji-yong, ZHAI Dong-jin, JIN Mu-hong

Advisor: LEI Hong

(Department of Mathematics and Computer Science, Tongren University, Tongren, Guizhou 554300)

**Abstract:** Aiming at solving the problem to analyze the intrinsic characteristics of heavy metal pollution in the urban topsoil, this paper studies the spatial distribution of the element concentration, the main causes underlying heavy metal pollution and the propagating character of the pollution. First, based on the sample data, we obtain the concentration distribution of each element by using Kriging interpolation method. And the spatial distribution map is further drawn by virtue of the Surfer software. With reference to the background values of concentration, the degree exceeding the stipulated standards of the various elements is attained in a functional zone based on the average actual concentration by relative contamination rate. In addition, we assess the environmental quality with respect to the geo-accumulation index and provide the pollution levels of each heavy metal element in different functional zone. Moreover, the distribution difference of various elements is analyzed based on the concentration distribution map. Based on such environmental difference, it is evident that human activities are the main cause of heavy metal pollution. Second, we establish the convection model for describing the pollutant spreading by virtue of the propagating character of heavy metal pollutant. All possible locations of polluting sources are calculated reversely using the characteristic curves of space coordinates. After analyzing the intrinsic principle underlying the model and evaluating the experimental results, we put forward some further research orientations, such as the increase of the sample size, the improvement of the interpolation algorithm, and the establishment of the time series evolution model provide some new ideas.

**Keywords:** heavy metal pollution, the index of Geo-accumulation, Kriging interpolation, convection model



假设A区的交通网络与平台设置的不意图中A路口 $i$ 的坐标为 $(x_1, y_1)$ , B路口 $j$ 的坐标为 $(x_2, y_2)$ 且A、B之间有道路连通,根据给出的比例尺,两路口距离即为

$$d_{AB} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \times 0.1 \text{ km}, \quad (1)$$

2) 利用Floyd算法<sup>[1]</sup>求解出每两个路口之间的最小路程

假设图G权的邻接矩阵为 $A_0$ ,经过计算,我们发现各交通警服务平台都无法在3min内到达的路口只有28,29,38,39,61,92这六个路口。

#### 4.1.2 模型约束

1) 0-1变量约束  $f_{ij}$ 为表示第 $i$ 路口 $i$ 的平台是否管辖第 $j$ 路口 $j$ 的0-1变量,即

$$f_{ij} \in \{0, 1\}. \quad (2)$$

2) 每个路,都必须被管辖的约束 要求每一个路口 $i$ 由且仅由一个平台管辖,即

$$\sum_{i=1}^{20} f_{ij} = 1. \quad (3)$$

3) 设置了平台的路口 $i$ 的约束 对于有平台的路口 $i$ ,应直接由该平台进行管辖 不应考虑其他情况,即

$$f_{ii} = 1. \quad (4)$$

4) 尽量3min内赶到事发地的约束 除了6个特殊的点以外,其他的点都要满足3min内有交通警赶到,即到交通警服务平台的距离应小于等于3km,因此有

$$f_{ij} \times d_{ij} \leq 3. \quad (5)$$

5) 对于6个3min内不能到达的路口,应该由与其距离最近的路口的平台来管理,即

$$d_{ij} \times f_{ij} \leq \min_i(d_{ij}). \quad (6)$$

#### 4.1.3 模型建立

我们建立各交通警服务平台工作量方差最小 乘法规划模型

$$\min \frac{1}{20} \sum_{i=1}^{20} (g_i - \bar{g})^2 \quad (7)$$

$$\text{s.t.} \begin{cases} g_i = \sum_{j=1}^{92} (f_{ij} \times n_j), \\ \bar{g} = \frac{\sum_{i=1}^{92} n_j}{20}, \\ \sum_{i=1}^{20} f_{ij} = 1, \\ f_{ii} = 1, \\ f_{ij} \times d_{ij} \leq 3, \quad j \neq 28, 29, 38, 39, 61, 92, \\ f_{ij} \times d_{ij} < \min_i(d_{ij}), \quad j = 28, 29, 38, 39, 61, 92, \\ f_{ij} \in \{0, 1\}, \\ i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 92. \end{cases} \quad (8)$$

#### 4.1.4 模型求解

我们先通过 Matlab 编程利用 Floyd 算法<sup>[4]</sup>, 求解得到任意两点之间的最短路程矩阵; 然后利用 Lingo 优化编程得到各交巡警服务平台的管辖范围. 计算得到这种划分方案各个交巡警服务平台管辖地区发案率的方差为 3.02.

#### 4.2 模型二 交巡警服务平台警力封锁调度问题

##### 1) 模型一建立

我们考虑到封锁路口时, 不仅要最后到位的平台时间尽可能短, 而且要求其他平台的速度也要尽可能的快. 即各平台到位的时间总和最小. 因此, 在第一层模型的基础上, 第二层规划模型将目标函数改为各个出警封锁的平台所用的时间总和最小.

第一层目标函数:

$$\min \max \left( f_{ij} \times \frac{d_{ij}}{v} \right). \quad (9)$$

第二层目标函数

$$\min \sum_{i=1}^{20} \sum_j f_{ij} \times \frac{d_{ij}}{v}. \quad (10)$$

约束条件:

$$\text{s.t.} \begin{cases} \sum_{i=1}^{20} f_{ij} = 1, \\ \sum_j f_{ij} \leq 1, \\ f_{ij} \in \{0, 1\}, \\ i = 1, 2, \dots, 20, \quad j = 12, 14, 16, 21, 22, 23, 24, 28, 29, 30, 38, 48, 62 \end{cases} \quad (11)$$

##### 2) 模型一的求解

通过计算, 按照上述方案能够在 8 02min 内, 以最快速度完成对所有进过 A × B 的交通要道路口的封锁.

#### 4.3 模型三 新增交巡警平台布置问题

##### 1) 解题思路及建模准备

经过上述计算我们得到交巡警 3min 内不能够到达 (即大于 3km) 的路口有 28, 29, 38, 39, 61, 92 号 6 个路口. 其中能够满足到 28, 29 号路口的路程小于等于 3km 的路口, 为 28 号和 29 号路口. 满足到 38, 39 号路口的路程小于等于 3km 的路口有 38, 39, 40 号路口. 满足到达 61 号路口好路口的路程小于等于 3km 的路口有 48, 61 号路口. 满足到达 92 号路口的路程小于等于 3km 的路口有 87, 88, 89, 90, 91, 92 号路口. 因此, 我们把不能满足出警时间大于 3min 的点归为 4 组, 在每组内, 要再新增一个交巡警服务平台, 就能够满足出警时间均小于等于 3 分钟的要求, 所以至少需要再新增 4 个交巡警服务平台.

##### 2) 模型三的建立

这里我们仍然采用多目标规划的方法, 在工作量最大平台的工作量尽可能小的前提下, 使工作量的分配尽可能得平衡.

第一层目标函数

$$\min \max \left( \sum_{j=1}^{92} f_{ij} \times n_j \right). \quad (12)$$

## 第二层目标函数

$$\min \frac{1}{24} \sum_{i=1}^{24} (p_i - \bar{p})^2. \quad (13)$$

## 约束条件

$$\text{s.t.} \begin{cases} f_{ij} \in \{0, 1\}, \\ d_{ij} \times f_{ij} \leq 3, \\ \sum_{i=1}^{24} f_{ij} = 1, \\ i = 1, 2, \dots, 24 \\ j = 1, 2, \dots, 92 \end{cases} \quad (14)$$

## 3) 模型一的求解

首先我们用Lingo编程求解第二层规划模型的结果为

$$\max \left( \sum_{j=1}^{92} f_{ij} \times n_j \right) = 8.5.$$

然后,将这个结果作为约束,代入第一层规划模型中,再用Lingo编程求解,得到各交巡警服务平台的管辖范围,其中A21平台配置在第28号路口或第29号路口, A22平台配置在第40号路口, A23平台配置在第48号路口, A24号平台配置在第87号路口。

这种分配方案各交巡警服务平台管辖地区发案率的方差为2.20,比没有新增平台之前的3.02相比有一定的改进,但部分分配更加趋于平衡。

## 4) 模型一的改进

从上面的模型计算结果我们可以看出,新增4个交巡警服务平台的方案虽然对各平台工作量很平衡有一定的促进作用,但平衡作用并不是很明显,最大工作量仍为8.5,此时我们应考虑新增5个交巡警服务平台的方案。

经过建模、编程、求解,我们决定将第5个新增的平台A25设置在第22号路口,这种分配方案各交巡警服务平台管辖地区发案率的方差为1.51,最大工作量为6.1,明显优于新增4个平台的结果2.20和8.5,说明各交巡警服务平台的平衡性得到显著的提升。

## 4.4 模型四 全市现有交巡警服务平台设置合理性及改进问题

### 4.4.1 对全市现有交巡警服务平台设置合理性评价

- 1 服务平台警力大都集中在A区,警力充足,而D区和F区的警力明显不足。
- 2 C区和E区每个平台处理的案件数较多,服务平台的工作强度较大。
- 3 D区交巡警服务平台管辖的面积大,人口多,而平台数却相对较少。
- 4 我们发现除了A区各平台3min内都无法到达的路口较少外,其他各区到每个平台的距离都大于3km的路口很多, B区有6个, C区有43个, D区有12个, E区有30个, F区有28个,这就会造成到某些路口的出警时间过长,无法满足迅速到达案发地点的要求。

### 4.4.2 全市警力重新配置模型的建立

#### 1) 建模准备

我们采用加权的方法来重新分配每个区的平台数量。我们确定影响交巡警服务平台分布的因素上要有各区路口总数,各区人口数和各区面积数。

我们采用极值标准化的方法对附表中给出的数据进行标准化,我们将标准化后的数值与对应的权值相乘,得到加权求和后的结果为: A区为0.02634, B区为0.38877, C区为0.842281, D区为0.592818, E区为0.807227, F区为0.69527

将上述数据换算成分比为: A区占15.9%, B区占9.8%, C区占21.3%, D区占15%, E区占20.4%, F区占17.6%。由于警务平台总数为80,所以重新分配的警务平台数依次为13,8,17,12,16,14。

## 2) 具体规划模型的建立

对每个城×交警平台布置位置的确定以及每个平台具体管辖范围的划分,我们是使用0-1规划模型进行最优方案的选取。下面我们以B区交巡警服务平台的重新规划为例进行规划模型的建立。

目标函数

$$\min \max \left( \sum_{j=93}^{165} f_{ij} \times \frac{d_{ij}}{v} \right). \quad (15)$$

约束条件:

$$\text{s.t.} \begin{cases} f_{ij} \in \{0, 1\}, \\ k_i \in \{0, 1\}, \\ \sum_{i=93}^{165} k_i = 8, \\ \sum_{i=93}^{165} f_{ij} = 1, \\ \text{若 } \sum_{j=93}^{165} f_{ij} \neq 0, \text{ 则 } k_i = 1, \\ f_{ii} = k_i, \\ j = 93, 94, \dots, 165, \\ j = 93, 94, \dots, 165. \end{cases} \quad (16)$$

## 3) 模型四的求解

我们利用Lingo软件求解,求解出交巡警服务平台在各区分布,如表1所示。

## 4.5 模型五的建立

### 1) 模型五的准备

由于我们知道犯罪嫌疑人逃跑的起点和时间,并假设其车速 $v$ ,警车速度和 $v_1$ ,则我们可以确定犯罪嫌疑人车辆在某一时间可能出现的地区范围,发现最小的包围圈为犯罪嫌疑人的目标。我们规定满足下列条件,就可认为交巡警能成功围堵住犯罪嫌疑人的车辆。

$$\frac{d_0(s)}{v} \leq \frac{d_f(s)}{v} - \Delta t, \quad (17)$$

其中 $d_0(s)$ 表示警务平台到s路口的最短路程; $d_f(s)$ 表示犯罪分子到s路口的最短路程; $\Delta t = 3\text{min}$ 。

通过对个体数据的分析,我们将不能够围堵住犯罪嫌疑人的路口全部选出来作为最初计划围堵的范围,这些路口为7,8,9,30,31,32,33,34,35,36,39,45,46,47,48,61。在这里,我们引入集合A、B(A、B集合之间均无重复的元素):

$$A = \{A_1, A_2, \dots, A_k\}, \quad p \in A, \quad B = \{B_1, B_2, \dots, B_k\}, \quad q \in B \quad (18)$$

其中  $A_i$  为集合  $A$  的子集, 表示在犯罪嫌疑人车辆路过第  $i$  个路口之后, 可能选择未被警力封锁的路口编号组成的集合,  $B_i$  为集合  $B$  的子集, 表示在犯罪嫌疑人车辆路过第  $i$  个路口之前, 可能选择的被警力封锁的路口的编号组成的集合。集合  $A$ 、 $B$  最后一个子集的编号, 均为  $K$ 。

表1: 警力重新分配后各区平台分布表

全市=城区	分配平台数	重新分配平台的位置/路口编号
A	13	1, 3, 5, 10, 11, 12, 13, 14, 29, 31, 40, 60, 91
B	8	94, 95, 97, 98, 100, 103, 125, 152
C	17	167, 168, 169, 171, 174, 175, 176, 177, 178, 180, 181, 200, 210, 239, 262, 275, 300
D	12	327, 328, 331, 332, 339, 343, 344, 360, 362, 363, 368, 370
E	16	372, 373, 374, 377, 379, 381, 383, 384, 385, 387, 388, 390, 396, 404, 419, 457
F	11	477, 478, 479, 480, 482, 483, 484, 485, 505, 541, 549, 574, 579, 582

## 2) 模型五的建立

目标函数

$$\min \text{account}(A), \quad (19)$$

约束条件:

$$\begin{cases} A_k = \emptyset \\ \sum_{i \in A} f_{iq} = 1 \\ \sum_{q \in B} f_{iq} = 1 \\ d_{iq} < d_{i+1,q} - \Delta t \\ i = 1, 2, \dots, 485. \end{cases} \quad (20)$$

## 3) 模型五的求解

通过求解, 我们求得需要派遣 22 个平台进行围堵,  $\text{account}(A) = 45$ , 其中最大的  $d_{iq}$  为 8.79 km, 即能够在 8.79min 内实现对犯罪嫌疑人车辆的围堵, 此时犯罪嫌疑人的车辆最多经过了 45 个路口。

## 5 模型评价

由于题目信息有限, 所以本文模型在实际应用时仍存在以下不足:

1) 在最后一问的求解中, 我们假设犯罪嫌疑人逃跑时车速为 60km/h, 但如果其车速加快, 则不能根据上述方法求解。

2) 对犯罪嫌疑人包围的方案只涉及到了包围, 而没有考虑如何抓捕, 可能会使得抓捕的时间比较长。



## 参考文献

- 1 韩卫华. 数学建模——方法[M]. 北京: 高等教育出版社, 2009.  
Han W H. Mathematical Modeling Method and its Application[M]. Beijing: Higher Education Press, 2009.
- 2 宋世忠, 贾东水. Matlab数值分析[M]. 北京: 机械工业出版社, 2009.  
Song Y Z, Jia D Y. Matlab Numerical Analysis and Application[M]. Beijing: Mechanical Industry Press, 2009.

## Research on Setting and Scheduling Scheme of Traffic and Patrol Police Service Platform

YE Yun-jia, LIU Jian, WANG Yu  
Advisor: CAO Hua-lin

Aeronautical Department of Mechanical Engineering, Naval Aeronautical  
Engineering Institute, Qingdao Campus, Qingdao 266111

**Abstract.** The paper investigates the distribution situation of the traffic and patrol police platform in a certain city. We develop a reasonable programming for traffic and patrol police service platform that needs to raise the resource utilization rate and reduce its response speed by using the Floyd algorithm in graph theory as well as the least square and multiple objective programming. Furthermore, we investigate the unreasonable factors underlying the current police scheduling condition and by combining with the practical situation, we further propose the measures, scheme that re-schedules the police strength. Besides, we establish the scheme of blocking up crossing roads based on specific situations.

**Keywords:** Floyd algorithm; least square method; multiple objective programming

文章编号 1005-3085(2011)07-0105-12

## 交巡警服务平台的设置与调度模型

但 琦<sup>1</sup>, 韩中庚<sup>2</sup>, 杨廷鸿<sup>1</sup>

1. 解放军后勤工程学院基础部 重庆 401311, 2. 解放军信息工程大学 信息学院 郑州 450002

**摘 要** 本文针对2011年全国大学生数学建模竞赛的B题“交巡警服务平台的设置与调度”问题,首先介绍了问题的背景与立意,然后分析给出了建立三个具体调度模型,二是具体模型,三是问题的主要求解方法和结果,接着简要分析,最后给出普遍存在的主要问题,最后提出了值得进一步研究的几个问题.

**关键词** 交巡警服务平台 交通网络图 优化模型 调度模型

**分类号** AMS:2000 30K45

**中图分类号** O221.2

**文献标识码** A

### 1 问题的背景与立意

“交巡警服务平台的设置与调度”问题作为2011年全国大学生数学建模竞赛本科组的B题,主要是根据重庆市交巡警平台设置的实际情况抽象加工而成的.

早在2010年2月7日,重庆市公安局交通巡警总队挂牌成立,之后部署了300个交巡警平台,8300多警力,担负起“破案打击、治安防控、交通管理、服务群众”四大职能.重庆市的交巡警服务平台,看上去就是一个占地约10平方米、高7米的“小亭子”.每个平台配备有1辆巡逻警车、4辆摩托车、GIS/GPS定位系统、视频监控镜头、移动查询执法终端、360度循环录像、高空定位爆闪灯、施救设施、防爆器材.集众多先进警务技术于一体.平台通过3G网络及专线与“大情报”平台、警务综合应用平台、社会信息资源平台和视频监控无缝对接,可随时查询比对人、车和车辆信息.平台还配备冰箱、微波炉、冷暖设备、安全警戒线和防暴设施,民警配备防暴靴、发炮背心,兼顾了人性化、安全性和舒适性.

交巡警不仅在形式上实现了110与122的合警联动,更建立了高效的指挥机制.警力接处警速度从过去的5到10分钟缩短为3到5分钟.在1城×任何一地点,能够在3小时内集结5台警车、20台警用摩托车、60名警力,实现对警情“第一时间、第一现场、第一信息、第一处置”.

交巡警平台不但能方便群众、缩短警民距离,也能预防犯罪、缩小犯罪空间.它替代了原有的上下个临街一道、老式破旧的治安岗亭和警务室,布设了案件高发、防范薄弱的重点部位.依托“大情报”系统,能紧盯刑事案高发时段、地段和交通管理乱点、堵点、事故多发点,做到警力跟着警情走,勤务件同民意行.交巡警制度整合了警力资源,构成了刑事执法、治安治理、交通管理、服务群众四大职能有机融合的新型防控体系.为了更有效地贯彻实施这些职能,需要在市区的一些交通要道和重要部位设置交巡警服务平台.由于警力资源是有限的,如何根据城市的实际情况与需求合理地设置交巡警服务平台、分配各平台的管辖范围、调度警力资源,是警务部门面临的一个实际课题.

这个问题我们最早是2010年4月在重庆的交巡警平台首次注意到,当时觉得交巡警平台很神秘,出于好奇提出了一些问题.例如,我们发现交巡警平台有的相距较远,有的相距较近,那么它们的位置是根据什么来选择确定的?一个平台分管多少地方?各平台之间是如何分工,又如何合作的?发生突发事件后是如何调度的?带着这些问题,当2010年8月再次到重

庆开会时,我们专门到交巡警平台去调研。此后又走访了交巡警总队进行咨询,收集了大量的相关数据资料和重庆市最新的多种版本地图等。直到2010年底讨论形成了一个初步的问题,并给出了问题解决的思路。后来经过全国命题研究会。根据专家组的意见,对题目又进行进一步的加工,直到2011年5月完成了基本问题的简要解答。此次又与专家组和广大专家多次交流修改和完善,直到竞赛最终完整地确定了这个题目。

## 2 A区交巡警服务平台的相关问题

这部分针对A区范围内的交巡警平台研究解决下面一个问题。

1) 要求合理地地为各个交巡警服务平台分配管辖范围,尽量在3分钟内平台交巡警能够到达各管辖的路口;

2) 若需要对进出该区的13条交通要道实施封锁,则要求给一个部的平台调度方案;

3) 根据现有的资料和数据,确定需要在该区增设平台的个数和具体设置的设置方案。

### 2.1 A区平台的管辖范围的合理分配问题

#### 2.1.1 A区交通网络赋权图和最短路径矩阵

将A区的交通线路抽象为交通网络赋权图。用 $x_i$ 表示第 $i$  ( $i=1,2,\dots,m$ )个路口, $y_j$ 表示第 $j$  ( $j=1,2,\dots,n$ )个交巡警服务平台。以路口 $x_i$ 为节点,路口之间的公路为边,其公路的长为 $d_{ij}$ 或边的权值, $d_{ij}$ 是赋权以建立A区的一个交通网络赋权图。将相应的邻接矩阵记为 $L=(l_{ij})_{m \times m}$ ,根据网络优化中求最短路径问题的Floyd算法<sup>[2]</sup>,用Matlab编程计算,任意两个节点 $x_i$ 之间的最短距离, $d_{ij}$ 和 $w_{ij}$ 的最短路径矩阵为 $A_{ij}$  ( $i,j=1,2,\dots,m$ ),单位:km。

#### 2.1.2 A区平台管辖范围的优化模型

要确定A区各平台的管辖范围,就是将A区内的每一个交通路口合理地分配给一个担任平台管辖的方案。这里所说的合理性主要是体现在两个方面。

1) 从能在3分钟内平台交巡警能够到达各管辖的路口,即要求每个平台到达所有管辖路口的最大时间尽量小;

2) 从实际出发,要求各平台的出警工作量尽量均衡。

为此,构造决策矩阵 $X=(x_{ij})_{m \times n}$ ,其中决策变量为

$$x_{ij} = \begin{cases} 1, & \text{若路口 } x_i \text{ 由平台 } y_j \text{ 管辖,} \\ 0, & \text{其它.} \end{cases}$$

用 $B_{ij}$ 表示A区内路口 $x_i$ 到平台 $y_j$  ( $i=1,2,\dots,m$ ;  $j=1,2,\dots,n$ )的最短路程(或时间),即 $B=(B_{ij})_{m \times n}$ 。在决策矩阵 $X$ 下,该区内各个路口到达其管辖平台的最短路程(或时间)矩阵为

$$T=(T_{ij})_{m \times n}=(X_{ij} \cdot B_{ij})_{m \times n}.$$

于是,对于平台 $y_j$ 来说,最大出警时间为

$$\max_{1 \leq i \leq m} T_{ij}, \quad j=1,2,\dots,n$$

另一方面,将A区内各个路口发案量 $w_i$ 为向量 $W=(w_1, w_2, \dots, w_m)$ ,其中 $w_i$ 表示A区内路口 $x_i$  ( $i=1,2,\dots,m$ )发案量,则各平台的工工作量可表示为 $G=W \cdot X=(G_1, G_2, \dots, G_n)$ 。

为了使各平台的工作量尽量均衡,则要求各平台工作量的标准差最小,即要求

$$\sigma(G) = \sqrt{\frac{1}{n} \sum_{j=1}^n (G_j - \bar{G})^2}$$

的最小,其中 $\bar{G}$ 为平均工作量

综上所述,以A区内所有平台的最大出警时间的最小和各个交警服务平台工作量标准差最小为目标函数,建立各平台管辖范围分配的双目标优化模型如下

$$\begin{aligned} \min_x \quad & \max_{1 \leq i \leq m} T_{ij}, \quad \min_x \sigma(G), \\ \text{s.t.} \quad & \begin{cases} \sum_{j=1}^n X_{ij} = 1, \quad i = 1, 2, \dots, m \\ \sum_{i=1}^m X_{ij} \geq 1, \quad j = 1, 2, \dots, n, \\ X_{ij} = 0 \text{ 或 } 1, \text{ 且当 } i=j \text{ 时, } X_{ij} = 1, \\ i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n. \end{cases} \end{aligned}$$

其第一个约束条件为每个路口只能有一个平台对其管辖,第二个约束为每个平台至少管辖一个路口及附近区域,第三个约束为决策变量,且平台所在的路口由该平台管辖

### 2.1.3 模型的求解与结果分析

对于上述的双目标0-1规划模型,直接求解是比较困难的。实际上,通过直观分析不难看出,有6个路口,在3分钟内是无法达到的,即为28与29、38、39、61和92号路口。为此,按照近原则,分别直接分配给15号、16号、2号、7号和20号平台管辖。除此之外的路口,都以满足在3分钟内到达的要求,即可将约束条件

$$\max_{1 \leq i \leq m} T_{ij} \leq 3$$

加入模型中。则将其转化为以各平台工作量均衡指标(标准差)最小为目标的0-1规划模型

另一种方法,将两个目标作线性加权和处理为单目标问题求解。在排除了6个3分钟内不能到达的路口后,最大出警时间就不再是主要的目标了,即对最大出警时间目标赋以较小的权值,而对工作量均衡目标赋以较大的权值。且Lingo软件直接求解,或用Matlab编程求解都可以得到相应的结果。求解方法不同,所取权值不同,其求解结果略有差别。

经可行分配方案的结果:20个平台平均最大出警时间为2.6759分钟,最大时间为5.7分钟,平均每天的出警次数为6.225次,最多的为11.5次,最少的为1.6次,不能在3分钟内到达的有6个路口。

## 2.2 A区20个平台对13个进出口路口的全封锁模型

从20个平台中选择13个对所要求封锁的目标路进行一对一的封锁。用 $(d_{ij})_{20 \times 13}$ 表示A区内平台 $y_i$  ( $i = 1, 2, \dots, 20$ )到目标路 $z_j$  ( $j = 1, 2, \dots, 13$ )的最短路(时间)矩阵。平台 $y_i$ 对目标路 $z_j$ 进行封锁的决策矩阵为 $X = (x_{ij})_{20 \times 13}$ ,其中当平台 $y_i$ 封锁路口 $z_j$ 时, $x_{ij} = 1$ ,其 $\bar{x}_{ij} = 0$ 。

在决策矩阵 $X$ 下,各平台到达要封锁目标路口的最短路程(时间)矩阵为 $T = (T_{ij})_{20 \times 13}$ , $d_{ij} \times x_{ij}$ ,完成封锁的最大时间为 $\max_{1 \leq i \leq 20} T_{ij}$ ,要实现全封锁的目标是最大时间最小,则

题归结为一个不完全的指派问题<sup>[2]</sup>, 其优化模型为

$$\begin{aligned} \min Y &= \max_{1 \leq i \leq 20} T_{ij} \\ \text{s.t. } &\begin{cases} \sum_{j=1}^{13} x_{ij} \leq 1, & i = 1, 2, \dots, 20. \\ \sum_{i=1}^{20} x_{ij} = 1, & j = 1, 2, \dots, 13. \\ x_{ij} = 0 \text{ 或 } 1, \text{ 且 } x_{ji} = 1, \\ i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 13. \end{cases} \end{aligned}$$

其中第一组约束表示一个平台最多只能封锁一个目标路口, 第二组约束表示每个目标路口只能由一个平台封锁. 用 Lingo 软件直接求解, 可以得到最优的结果. 最优的封锁调度方案和相应的封锁时间, 如表 1 所示.

表 1 A 区的全封锁调度方案 (时间单位: 分钟)

平台	封锁路口	封锁时间	平台	封锁路口	封锁时间	平台	封锁路口	封锁时间
A2	38	3.9822	A9	16	1.5325	A14	21	3.265
A4	62	0.35	A10	22	7.7079	A15	28	4.7518
A5	18	2.4758	A11	24	3.8053	A16	14	6.747
A7	29	8.155	A12	12	0			
A8	4	1.0608	A13	23	0.5			

要观测个封锁的最长时间为 8.0155 分钟, 总平均封锁时间为 3.55296 分钟. 实际上, 这个最优的封锁方案应该是唯一的.

### 2.3 A 区需要增加平台的设置模型

在实际上, 要新增设一个平台必然会需要一定的建设和运行成本, 将这些成本视为投入. 新增平台后, 将需要对全区各平台的管辖范围进行重新分配, 从而会使得所有平台在辖区内的最大出警时间和相应工作量指标发生改变, 将这种改变的效果视为产出. 那么这个问题就可以视为一个“投入与产出”的问题.

在现有平台的基础上, 因为设有设置平台的路口节点有  $m-n$  个, 即考虑在这  $m-n$  个可能的节点中任取一个增设平台. 按照问题 1) 的方法, 对增加平台后各平台的管辖范围进行重新分配. 则一定存在一个增设平台的方案, 使得“产出”有最大值. 此方案就是增设一个平台的最优方案. 事实上, 只要给出增设一个平台的模型和求解方法, 则增设两个、三个、四个和  $j$  个的情况可以同理可得.

对没有设平台的路口进行重新排序, 并记为  $x_1, x_2, \dots, x_{m-n}$ , 则可能要增设一个平台位置的决策变量记为  $r_1, r_2, \dots, r_{m-n}$ , 其中若路口  $x_i$  设置新平台  $r_i = 1$ , 其他  $r_i = 0$ .

考虑增加一个平台. 则平台的个数由原来的  $n$  个变为  $n+1$ , 由问题 1) 的方法, 对  $n+1$  个平台的管辖范围进行重新分配, 并计算所有平台的最大出警时间, 记为  $t_1$ , 相应的, 工作量标准差为  $\sigma_1$ . 记增设平台后所有平台的最大出警时间为  $t_0$  和工作量的标准差为  $\sigma_0$ . 定义增设一个平台后的产出值 (增设效益) 为

$$F = \rho \frac{t_0 - t_1}{t_0} + (1 - \rho) \frac{\sigma_0 - \sigma_1}{\sigma_0},$$

其中  $\rho$  为经验比例系数。于是, 可得如下增设平台的优化模型:

$$\begin{aligned} \max F &= \rho \frac{t_0 - t_1}{t_0} + (1 - \rho) \frac{\sigma_0 - \sigma_1}{\sigma_0} \\ s.t. &\begin{cases} \sum_{i=1}^m \sum_{j=1}^n r_{ij} = 1 \\ r_{ij} = 0 \text{ 或 } 1, \quad i = 1, 2, \dots, m - n. \end{cases} \end{aligned}$$

对于模型的求解, 参照问题 (1) 的方法, 便可以得到增设一个平台的位置和产生的效益值。在考虑增设第一个平台之后, 再考虑增设第二个平台, 依此类推。不妨取  $\rho = 0.5$ , 讨论增设 1 个至 5 个平台后所得到的效益情况比较结果, 如表 2 所示。

表 2 增设新平台后对应效果 (时间单位: 分钟)

增设平台情况	出警时间	新位置	产出效益	增设平台情况	出警时间	新位置	产出效益
无增设平台	5.7105	2.6000		增设 3 个平台	3.6822	2.4157	0.145802
增设 1 个平台	4.1902	2.9545	0.077239	增设 4 个平台	3.6822	1.9862	0.086828
增设 2 个平台	3.6822	3.6593	0.110575	增设 5 个平台	3.6071	8.426	0.3807

从表 2 可以看出, 增设 4 个平台时效果较好, 最大出警时间变化不大, 作业标准与表 1 比较小, 具体的增设位置为 29、10、18 和 87 号路口。

### 3 全市交巡警服务平台的相关问题

这部分针对全市主城区区的交巡警服务平台需要解决下面两个问题:

1. 要求评价该市区设置平台方案的合理性, 如果明显不合理, 也给出解决方案;
2. 该主城区已发生 1 起重大刑事案件, 要求给出满足全市平台警力资源的最佳布防方案。

#### 3.1 全市平台设置方案的合理性评价模型

该问题可以从两个方面考虑。一是全市六区的平台不考虑分区限制, 即将全市六区视为一个整体的网络, 予以统一调度管理。二是全市六区的平台分区调度管理, 即分别考虑六个区的具体情况。在这里, 考虑第一种情况, 第二种情况与问题 (1) 的方法相类似。

为, 全市平台设置合理性的主要指标有最大出警时间

$$T = \max_{1 \leq j \leq m} (t_j),$$

其中  $t_j, j = 1, 2, \dots, n$  表示平台  $j$  管辖区域内的最大出警时间。相应的作业量  $G_j, j = 1, 2, \dots, n$ , 其均值为  $\bar{G}$ , 最大作业量为

$$G = \max_{1 \leq j \leq n} (G_j),$$

则各平台的工作量标准差为

$$\sigma = \sqrt{\frac{1}{n} \sum_{j=1}^n (G_j - \bar{G})^2}.$$

根据问题附件中所给全市的相关数据和A区平台管辖范围的分配模型,用启发式算法求解,她可以得到全市平台管辖范围的分配方案.从全市情况来看,在全市的582个路口节点中,总共有138个是不能在3分钟内到达的,按就近原则分配这些路口节点.将其余的路口节点在到达时间不超过3分钟的约束条件下寻求最优的分配方案,不难得到全市的80个平台的平均最大出警时间为3.7735分钟.最大的出警时间为12.6803分钟.平均每天的出警量为84312次.最多的出警量为261次,最少为16次.由此可知,现有全市的平台设置还不尽合理,需要再增设一些新的平台.

3.2 增设新平台方案的确定模型

交巡警平台设置为合理性主要体现在两个方面.各平台的最大出警时间尽量小和总工作量尽量均衡.解决该问题的方法与A×的解决方法不同.因为路口数量多,不能对所有路,进行全搜索实现,要根据现有平台设置明显不合理的地方进行增设新平台.因此,在这里定义一个与平台的出警时间和工作有关的指标,称为“需求度”,作为判断在一个路口是否需要增设平台的依据,从而找出最需要增设平台的路口.对于每个路口*x<sub>i</sub>*,定义需求度如下:

$$\mu_i = \lambda \frac{t_i}{t_{\max}} + (1 - \lambda) \frac{G_i}{G_{\max}},$$

其中

$$t_i = \sum_{j=1}^n T_{ij} = \sum_{j=1}^n B_{ij} X_{ij}$$

为路口*x<sub>i</sub>*到其所管辖平台的距离;

$$t_{\max} = \max_{1 \leq i \leq n} T_{ij}$$

为所有出警时间的最大值. *G<sub>i</sub>*为管辖路口*x<sub>i</sub>*平台的工作量,

$$G_{\max} = \max_{1 \leq i \leq m} (G_i)$$

为所有平台最大工作量.  $\lambda$ 为权重.对于每一个没有设平台的路口,计算其相应的需求度值,如果需求度较大的路口比较集中在一个区域,则说明在该区域现有平台的设置有明显的欠合理,从而可在该区域增设新平台来降低需求度.根据上述的方法,对比增设1个至6个平台的情况,如表3所示.

表3: 增设新平台的对应结果比较分析

增设平台数	增设平台编号	最大出警时间(分钟)	最大工作量(次/天)	工作量标准差
0	未增设	12.6803	26.1	4.4946
1	314	12.6803	18.4	4.1365
2	517	12.6803	18.4	3.9251
3	388	9.8992	18.4	3.3753
4	330	8.1188	15.9	3.9973
5	288	8.1188	15.9	3.8714
6	206	8.1188	15.9	3.8794

从表3可以看出,前5个平台增设后都对一个指标产生明显的改善效果,但是增加第6个平台后却没有对指标产生明显的变化。故在全市范围内增设5个平台就可以改善现有平台设置明显不合理的状况。

### 3.3 全市范围内最佳围堵方案的确定模型

将事发地点P到其各路路口节点的最近路程(或时间)用矩阵 $D = (D_1, D_2, \dots, D_m)$ 表示。不妨设交巡警和犯罪嫌疑人逃跑的车辆速度均为60 km/h。在实际中,犯罪嫌疑人逃跑的时间不同,可能逃跑的区域范围也不同。其逃跑的区域范围可以看作以事发地P为中心沿交通网络向外辐射的一个×域。在时间 $t$ 内将逃跑区域内所有路口节点的集合记为 $F(t)$ ,即为嫌疑人逃跑 $t$ 时间后可能逃跑的×域范围,则有 $F(t) = \{x_i | D_i \leq t, 1 \leq i \leq m\}$ 。要将犯罪嫌疑人封锁在 $F(t)$ 范围内, $F(t)$ 的所有边界路口节点集合记为 $E(t)$ ,即 $E(t)$ 中的任一节点都属于 $F(t)$ ,且与 $F(t)$ 之外的邻近节点邻接,故有

$$E(t) = \{x_i | x_i \in F(t), x_j \notin F(t), l_{ij} < \infty, 1 \leq i \leq m, 1 \leq j \leq n\}.$$

其中 $x_j \notin F(t)$ 为 $x_i \in F(t)$ 的×域外邻接点, $l_{ij}$ 为邻接距离。

对于任一给定的路口节点 $x_i \in F(t)$ ,需要封锁的时间为 $D_i - 3$ 分钟。因此要全封锁区域 $F(t)$ 即 $E(t)$ 内所有路口节点需要的最短时间为 $\min_{x_i \in E(t)} D_i - 3$ 分钟,最长时间为 $\max_{x_i \in E(t)} D_i - 3$ 分钟。

在全市范围内最优的封锁方案应该使得封锁×域尽量小,而封锁×域的大小可以由集合 $F(t)$ 的秩(节点个数)来表示,即 $Q = \text{Rank}(F(t))$ 。要保证封锁范围的有效性,即确保封锁的必要条件为 $B_{ij}X_{ij}(t) \leq D_i(t) - 3, i = 1, 2, \dots, p, j = 1, 2, \dots, n$ ,其中 $X_{ij}(t) = 0$ 或1,即表示 $t$ 时间要平台 $y_i$ 去封锁路口 $x_j$ 取1,否则取0, $p = \text{Rank}(E(t))$ 。则有最佳的全堵优化模型:

$$\min Q = \text{Rank}(F(t)).$$

$$\text{s.t.} \begin{cases} F(t) = \{x_i | D_i \leq t, 1 \leq i \leq m\}, \\ E(t) = \{x_i | x_i \in F(t), x_j \notin F(t), l_{ij} < \infty, 1 \leq i \leq m, 1 \leq j \leq n\} \\ B_{ij}X_{ij}(t) \leq D_i(t) - 3, i = 1, 2, \dots, p, j = 1, 2, \dots, n, \\ p = \text{Rank}(E(t)) \end{cases}$$

这里决策变量 $X_{ij}(t)$ 由下面的优化模型确定

$$\min Y = \max_{1 \leq i \leq p, 1 \leq j \leq n} B_{ij}X_{ij},$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^p X_{ij} \leq 1, j = 1, 2, \dots, n, \\ \sum_{j=1}^n X_{ij} = 1, x_i \in E(t), i = 1, 2, \dots, p, \\ \sum_{j=1}^n X_{ij} = 0, x_i \notin E(t), i = 1, 2, \dots, p, \\ x_{ij} = 0 \text{ 或 } 1, i = 1, 2, \dots, p, j = 1, 2, \dots, n. \end{cases}$$

对于上面的复杂的优化模型采用启发式搜索算法,可以得到一种有效的全封锁方案,具体结果如表4所示。



表4 种有效的全封锁方案和对应封锁时间(时间单位:分钟)

平台编号	封锁节点	封锁时间	最短逃跑时间	平台编号	封锁节点	封锁时间	最短逃跑时间
A16,16	16	0	3.30.6	C5(170)	243	5.0978	8.3009
A5,5	5	0	3.8768	F1(175)	561	4.3548	8.7969
A6,6	6	0	3.9074	A4,4	60	1.7792	8.8017
C8,73	235	0.5315	4.1282	C4(169)	240	7.0474	10.1541
A,5(15)	15	0	4.1388	A1(1)	41	4.4412	10.5025
A3,3	55	1.2654	5.2104	C3(168)	4	3.6904	10.7557
C6,17	232	1.9064	5.7845	C17(182)	273	2.1024	12.8088
A13,7	10	0	6.1882	C10,175	168	4.9779	13.3739
C17,172	244	1.5289	6.8421	D1(320)	370	7.8085	16.7721
A2,2	5	2.1117	7.1106	D2,321	371	8.0798	17.1179
A17,17	10	2.6879	7.9625	C2(167)	248	3.0788	17.5237

从表4可以看出,一旦封锁时间小于嫌疑人逃出封锁区域的最短时间减去3分钟时,即交警警能在嫌疑人之前到达要封锁的目标路口,则实现了对该路口有效封锁。需要对22个路口进行封锁。最长的封锁时间为8.9708分钟,即调度全市的交警警+11的警力,在9分钟之内可以完成有效全封锁。从交警需要封锁的时间和嫌疑人逃出区域的时间来看,该封锁方案是有效的,即能确保将嫌疑犯被封锁在该区域范围之内。

#### 4 关于竞赛论文的综合评述

交警警服务平台的设置与调度问题作为2011年“高教社杯”全国大学数学建模竞赛本科组的B题,根据在全国和部分赛区的企业统计情况,大约有一分之一的本科组参赛队选择了这个题目。这个题目的特性和解决难度,我们在赛前估计选做这个问题的人会远远预料之外的,但没想到做A题数量比如此之大出乎预料之外的,这成为命题人的一个遗憾。赛后,我们专门邀请导师交流和评委专家分析,出现这种情况的主要原因有两点,一是该题密切相关的参考资料几乎没有。二是解决这个问题对模型表述、模型的求解算法和编程能力要求都比较高,这让很多同学望而却步。

从另一个侧面也反映出来,我们的学生对文献资料的依赖性过强。而针对实际问题独立思考和创新能力却不够强。同时,对实际问题数据处理能力、算法设计、编程能力也不够强。总体反映对数学建模的能力不足。因此,如何加强培养学生独立思考和创新能力,提高数学建模能力,都是值得广大的数学建模指导教师思考和在培训中注意的问题。

尽管如此,我们也高兴地看到,很多参赛队在有限的时间内非常好地解决了这个复杂的问题。而且有的参赛队给出很好的解决思路和方法,都是很值得赞赏的。譬如,武汉大学的一个参赛队所建立的全市范围的围堵模型很有特色,建模思路清晰、模型表述简洁、自己设计的启发式去巧妙。程序实现也很有创新之处。为此获得了全国一等奖。但在文字表述和符号的使用上也有不足之处。上海财经大学一个参赛队的围堵模型也有不错的表现。另外,被选择发表的一篇优秀论文都有各自不同的特点,多数都是基本问题完成的比较好,论文表述也比较清晰,但对全市的围堵模型普遍完成的不好(具体可参考他们发表的论文)。当然,论文是学生在规定时间内完成的,存在一些差错和不足均属正常。

#### 4.1 解决问题有代表性的方法概述

第一部分, 针对A区的情况主要包括一个小问题, 首先需要根据题目所提供的数据, 计算由交通网络节点间的邻接矩阵和最短路径矩阵等, 这也是解决问题的基础, 然后按题目要求分别解决三个问题

1 针对为A×20个平台合理分配管辖范围的问题, 首先要确定合理性的指标。通常包括平台的出警时间尽量在3分钟之内和任务量尽量均衡, 然后以每一个平台在所管辖范围内的最大出警时间最小和各平台的任务量的方差(或标准差、绝对差等)为目标, 建立20个平台管辖范围的最优分配模型, 即归结为一个不完全的指派模型。可以直接求解该目标的优化模型, 也可以根据实际情况简化为单目标优化模型来求解。可以用Lingo直接实现, 也可以用Matlab编程实现求解。

2 针对A×13个出入口的全封锁问题, 以最长的封锁时间最小为目标, 以一个平台最多封锁一个路口和每一个路口有且仅有一个平台封锁为约束, 建立问题的指派模型, 直接用Lingo求解即可得到最优的封锁方案。

3 针对在A×增加2至5个平台的方案问题, 首先确定一个需要增加平台的原则和判断增加平台位置的准则, 判断在哪里最需要增加平台。建立增加2至5个平台后, 再按问题1的方法重新分配各平台的管辖范围, 计算各平台的最大出警时间和出警任务量的均衡性指标。根据增加平台后的效果, 确定出需增加平台的个数和位置。同时要说明需要增加k个, 而不需增加k+1( $2 \leq k \leq 4$ )个的理由。

第二部分, 针对全市六区的情况主要包括两个小问题, 分析评价全市现有平台设置的合理性和确定最佳的值班方案。

1) 针对分析评价全市六区现有平台设置的合理性问题, 可以在两种不同的假设下考虑问题。一种是将全市六区视为一个整体的网络图, 不考虑各区的外限。一个由40个平台可以跨×管辖邻近路, 另外, 参照A×平台管辖范围分配方法, 即可得到管辖范围分配方案是一个整体的评价方案。事实上, 在全市582个路口中只有138个是不可能3分钟内到达的路口, 因此, 按就近原则分配这些路口。将其它路口节点在到达时间不超过3分钟的约束条件下, 按各平台的任务量尽量均衡的原则就近分配管辖范围。另一种是将全市六区分×管理, 即分别视为6个独立的交通网络图, 采用A×分配管辖范围的方法, 分别给出各区的分配方案和合理性的评价结果。判断合理性的指标为各平台的最大出警时间和任务量的均衡性。对比, 明显不合理的地方增派新平台的方法与A×的情况类似。

2) 针对确定全市范围内围堵方案问题, 首先要确定在这段时间内嫌疑人可能逃跑的最大×域, 针对这个×域找出需要封锁的目标路口(节点)。再利用对A×全封锁问题的解法方法, 合理调度全市的80个平台的警力(不受区的限制), 对可能逃跑区域范围的一标路口进行全封锁即可。

事实上, 选择封锁的×域越大, ×域内所包含的路口节点就越多。实际要进行搜捕的工作量会越大; 实现封锁时间越长, 犯罪嫌疑人逃跑的范围可能就越大, 如果选择封锁的×域越小, 实现搜索的时间就越短, 在犯罪嫌疑人可能全跑出封锁区域。我们既要求合理地选择封锁范围, 从而保证安全封锁, 而同时又使得所需要的封锁时间尽量少。既要实现全封锁的时间最少, 又要封锁的×域最小。即×域内的路口节点数最少。

要求确定封锁方案, 就是以封锁区域最小, 也可以需要封锁×域内的路口节点数最小为目标。建立实现全封锁方案的优化模型。注意到, 需要封锁区域的大小是随时间变化的, 能封锁住的基本约束条件是

$$\text{出事地点到将要封锁的路口所需时间} + 3 \text{ 分钟} > \text{指派平台到封锁路口的所需时间}$$

对于问题的求解需要设计启发式算法编程实现,具体的实现需要一定的技巧和方法

## 4.2 竞赛论文存在的主要问题概述

浏览这个国和部分赛区评审的参赛论文,普遍存在的问题主要有以下几个方面

### 4.2.1 A区相关问题的情况

1) 针对A区为交巡警服务平台合理分配管辖范围的问题。很多论文都是依就近分配的原则来分配管辖路口节点,即以最短路径或最短出警时间为目标。将72个路口节点分配给20个平台。但没有从实际出发适当考虑各平台出警工作量的均衡性问题。导致了分配结果明显的不合理,甚至有几个平台只管平台所在的一个路口,典型的有6号、10号和14号平台,有的平台管的路口很多,工作量严重的失衡。虽然,问题中没有明确要求考虑工作的均衡,但从数学建模解决实际问题的角度,应该要考虑这个问题。这也反映在参赛学生对数学建模创新性的认识,理解还不够,更多的是基于应用题方面上就事论事地回答问题。

2) 在考虑对A区的13个路口实施紧急封锁的问题时,很多论文都是以封锁三成的总时间或平均时间最小为目标,建立指派模型,求解0-1规划模型得到封锁方案,这都是不太合理的。解决该问题的核心是用最短时间封锁所有的路口,所以应该取最大的封锁时间最小化,和 $hw$ 的讨论才是完成全封锁的时间。

3) 在考虑增设平台的问题时,很多论文只考虑3分钟内到达的要求目标,即以相对现在6个3分钟内不能到达的路口,通过直观分析就确定增设4个平台的结果。而没有兼地综合考虑各平台出警工作量的均衡的内涵。同时没有通过求解优化模型给以增设 $k(1 \leq k \leq 5)$ 个平台,而不需要增设 $k+1$ 个平台的理由,即分析比较增设2至5平台的最优选择。

### 4.2.2 全市六区相关问题的情况

1) 对于全市六区现有平台设置的合理性分析问题。很多论文把这个问题看成一个综合评价问题,用了简单的层次分析法、综合评价法、TOPSIS方法等。也简单地得到一个合理与不合理的结论。凡此种论文都没有充分地说明合理与不合理的理由,更没有给出解决不合理的具体的方法。如增设平台,增设多少个和增设的具体位置在哪里,增设以后的效果怎么样,如何重新分配管辖范围等问题。

2) 对于全市范围内交通路口拥堵的问题,绝大多数的论文都没有给出可行的拥堵方案。很多人把问题看成一个确定性的问题,有的直观分析选择要封锁的路口,就近确定平台执行封锁。还有的建立了一个确定的指派模型求解得到一个确定的方案。甚至,有不少论文没有考虑到保证封锁住的基本条件。所给出的方案没有办法保证能封锁住。还有个别的论文,简单地直接去封锁13个路口,没有考虑后续追捕问题。

3) 很多论文所给出的模型没有详细的分析与解释,让别人很难看懂。不仅仅有模型没有具体的求解算法,只是说用软件求解得到结果。有的也没有详细的程序。难易判断其模型和结果的准确性。有的论文所给出的结果只是些数据、或图表,没有任何分析与说明,人很难判断其有效性和正确性。

## 4.3 进一步研究的问题

虽然2011年全国大学生数学建模竞赛已经结束。但从我们评阅的论文来看,对于交巡警服务平台的设置与调度问题无论是从实际问题本身。还是从相关问题理论、模型和算法等方面,仍然还存在一些有待于进一步研究的问题。在这里提出来,希望有兴趣的同学和老师做进一步深入的研究。

1) 关于平台的管辖范围合理分配问题。是否能用区域覆盖的方法,给出一个连续区域分配方案,而不是仅考虑路口节点。

2) 如果全市范围内内的交巡警平台可以统一调度,没有区的界限,则能否给出全市所有平

台管辖范围的合理分配方案,要有相应的数学模型和相应有效的求解算法与结果

3) 从全局角度出发,不考虑区域界限。根据现有平台的设置情况,从110警时区、112短和各平台的110警工作力量均衡两个方面考虑,通过建模分析提出增加平台的总体方案,包括具体的数量和位置

4) 关于全市范围“动态”围堵数学模型简洁准确的表述方法和求解算法的进一步研究

5) 关于全市范围内围堵的最好方案究竟是什么? 需要围堵区域是什么? 包括哪些路口节点? 现有、确定的最好围堵区域结果包含21个到22个路口节点

6) 对现今110范围内的围堵最快的时间是多少? 现有、确定的最好的结果是89分钟。是否还有更快的可行方案?

7) 能否证明目前的结果已是最好的结果?

8) 根据现有围堵方案所给出的区域范围,进一步研究在这个区域内最有效的搜捕模型

对于上述问题的进一步研究都是有意义的,研究成果需要给出具体、准确的数学模型、求解算法、明确的结果和可靠性分析,必要情况需整给出理论证明和程序代码

## 5 结束语

2011年“高教社杯”全国大学生数学建模竞赛工作圆满结束了,也正是在迎来这项赛事20周年之际,我们能为竞赛活动以微薄之力,有幸成为今年的命题人,倍感荣幸。我们也欣喜地看到,2011年的竞赛规模达到了19190个队,并且多个国家和地区的11个境外参赛队参赛。即我们在竞赛上走中了日子,走向了世界。这也充分说明了这项活动的生命力对数学建模事业的号召力。在此,我们衷心地祝愿这项竞赛活动更健康有序地发展,进一步提高我国建模水平和应用水平,努力推动数学建模事业的发展,发扬数学建模的精神,为社会培养更多高素质的人才。

致谢 在题目形成和加工过程中,得到了全国组委会、各位专家的指导和帮助,特别是周义仓、姜启源、王长志和边锋四位教授多次在一起深入细致地讨论问题。谭永基教授和谢金荣教授等也都给予了很多的帮助。还有重庆市交巡警支队有关领导给予了很大的支持。在这里一并表示衷心的感谢。

## 参考文献:

- [1] 全国大学生数学建模竞赛组委会. 2011年竞赛题 OL. <http://www.mcm.edu.cn>. 2011  
Organizing Committee of the National Mathematical Contest in Modeling. 2011 CUMCM Problems OL. <http://www.mcm.edu.cn>, 2011
- [2] 韩俊. 数学建模方法及其应用(第2版)[M]. 北京: 高等教育出版社, 2009  
Han J G. Methods and Applications of Mathematical Modeling (2nd Edition)[M]. Beijing: Higher Education Press, 2009
- [3] 赵静, 但琦. 数学建模与数学实验(第3版)[M]. 北京: 高等教育出版社, 2008  
Zhao J, Dai Q. Mathematical Modeling and Mathematical Experiments. 3rd Edition [M]. Beijing: Higher Education Press, 2008

## Optimized Model of Setting and Scheduling of Traffic and Patrol Police Service Platforms

DAN Qi<sup>1</sup>, HAN Zhong-geng<sup>2</sup>, YANG Ting-bong<sup>1</sup>

Department of Basic Logistics Engineering University PLA Chongqing 401311

2- Institute of Information Engineering Information Engineering University PLA Zhengzhou 450002

**Abstract:** In this paper, according to the problem B of 2011 Higher Education Press Cup CUMCM, the background of the problem of setting and scheduling of traffic and patrol police service platforms is introduced. Then the approaches and the concrete models of several related problems are given, as well as the main solving methods and the corresponding results, and the general existing problems of the participating papers are analyzed briefly. Finally, several problems worthy of further study are given.

**Keywords:** traffic and patrol police service platforms; the transportation network map; model of distribution surrounds mode

文章编号 1005-3085 201107 0117-08

## 交巡警服务平台的设置与调度

胡立忠、张金磊、朱金长

指导教师: 李安平, 安萍力, 樊欣迪

(空军工程大学导弹学院, 西安 713800)

續書按 該文已略論陳 維理亦曾 表 示 願 意 移 往 科 林 托 拉 瓦 亞 亞 國 等 地 辦 事， 何 何 題 目 上 似 較 為 專 制， 在 提 煉 生 動， 除 那 個 題 目 不 變， 其 他 3 個 題 目 均 有 一 個 與 移 民 一 字 有 關， 其 中 第 二 個 題 目 是 不 同 題 的 因 此 在 食 料 上 用 了 3 個 不 同 的 字 眼 來 表 示， 而 第 三 個 字 眼 則 對 應 題 目 中 的 內 容。

[illegible]

关键词 Q-1 规划, Floyd 算法, 最短路径问题, 动态规划

分类号 AM5 (K3) (K2) (50) (52)

中图分类号 O221

文献标识码: A

## 1 问题分析

### 1.1 问题一的分析

1. 要为各交通警服务平台分配管辖范围。首先需要利用图论中的相关算法求得任意平台自身最短路径。其次,考虑依据就近原则将路点分配给距离最近的交警平台管辖,这样就可以最大限度满足管辖范围的分派要求。

2 要从 20 个交警警务服务中, 选择 13 个在道路拥堵时段, 合理的调度方案以满足 13 条主要道路, 最后一个到达交通要道路口的交警警务服务中, 所耗时间最少。如需要进一步考虑该方案的优化, 则还应满足出警总路程达到最短。

3 增加平台时可将 5 分钟内到达案发现场这一要求作为约束条件,求得不能满足要求的路口,为使这些路口满足条件,应在这些路口附近(包括本身)进行增设。考虑到工作量的均衡,可将工作量的平方作为衡量冗余度的指标,进而确定增加平台的个数与位置。

## 1.2 问题二的分析

1. 本问题要求按照设置交警服务平台的原则和任务, 分析研究该市现有交警服务平台设置方案的合理性。因此, 建立模型时, 综合考虑最大出警时间和 1 件量的均衡性这两个因素, 在所得结果的基础上再对不合理的平台提出改进方案。

2) 由于在案发后, 犯罪嫌疑人会立即逃窜, 且其逃窜路线是随机的, 对此, 在建立模型时, 考虑的是将“出事地点到将要封锁的路口所需时间大于等于指派平台到封锁路口的所需时间加3分钟”作为能封锁住的基本约束条件, 并在这个约束条件下给出最优封锁方案。

2 模型的建立与求解

在建立模型前，首先将警车实际行驶速度 $v=60\text{ km/h}$ 转换为图上行驶速度 $v'$ ，由地图比例尺计算公式得

$$v'=\frac{v}{100000}=10\text{ mm/min},$$

即出警时交巡警每走一分钟对应图上10mm的路程。其次，利用图论中Floyd算法通过Matlab软件<sup>[2]</sup>求出任意平台与任意路口间的最短距离。

2.1 问题一模型的建立与求解

2.1.1 管辖范围分配问题

为满足尽快快速到达事发地的要求，采用就近原则将各路 $l$ 分配给路程最近的服务平台管辖。记第 $i$ 个交巡警平台距第 $j$ 个交通要道路径为 $l_{ij}$ ，建立目标函数为

$$f_j=\min_{i=1}^{20}l_{ij},\quad j=1,2,\cdots,92$$

以体A区交巡警平台管辖范围的分配方案，如表1所示。

表1 A区交巡警平台管辖范围的分配方案

平台	管辖范围	平台	管辖范围	平台	管辖范围
1	1, 67, 68, 69, 71, 73, 74, 75, 76, 78	2	2, 39, 40, 43, 44, 70, 72	3	3, 54, 55, 65, 68
4	4, 57, 60, 62, 63, 64	5	5, 49, 50, 51, 52, 53, 56, 58, 59	6	6
7	7, 30, 32, 47, 48, 61	8	8, 33, 46	9	9, 31, 34, 35, 45
10	10	11	11, 26, 27	12	12, 25
13	13, 21, 22, 23, 24	14	14	15	15, 28, 29
16	16, 37, 38	17	17, 41, 42	18	18, 80, 81, 82, 83
19	19, 77, 79	20	20, 84, 85, 86, 87, 88, 89, 90, 91, 92		

需要说明的是，通过计算发现A区中编号分别为28, 29, 38, 39, 61, 92的6个路口交巡警无法在3分钟内到达。距离最近交巡警服务平台到达这6个路口的最短时间，如表2所示。依据就近原则，这6个路口只能分配给距它们最近的平台管辖。

表2 最近平台到达6个交叉路口的最短时间(单位：分钟)

平台	路口	最短时间	平台	路口	最短时间	平台	路口	最短时间
16	38	3.40586	7	61	4.1902	2	39	3.68219
20	92	3.60127	15	28	4.75184	15	29	5.70053

### 2.1.2 警力合理调度问题模型的建立与求解

#### 1) 最快全封锁模型的建立与求解

基于问题分析,可将题意表述如下:寻找一种方案,在20个交巡警服务平台中选择13个,对13条交通要道进行封锁,使封锁时间最小。为满足快速实现全封锁,采用以下步骤:

- 在20个交巡警服务平台中选择13个,对13条交通要道进行封锁,得到一个可行方案;
  - 每种可行方案对应13条出警路线,封锁时间取决于其中最长出警路线;
  - 对这些最长出警路线进行比较,选择其中的一个作为调度方案。
- 因此可得模型如下:

$$\min c = \max_{1 \leq j \leq 13} \left\{ \sum_{i=1}^{20} c_{ij} \right\}$$

$$\begin{cases} \sum_{i=1}^{20} a_{ij} = 1, \\ \sum_{j=1}^{13} a_{ij} \leq 1, \\ c_{ij} = a_{ij} l_{ij}, \end{cases}$$

其中

$$a_{ij} = \begin{cases} 1, & \text{第 } i \text{ 个平台封锁第 } j \text{ 个交通要道} \\ 0, & \text{第 } i \text{ 个平台不封锁第 } j \text{ 个交通要道} \end{cases}$$

$l_{ij}$  表示第  $i$  个交巡警平台到第  $j$  个交通要道的最短路程,  $c_{ij} = a_{ij} l_{ij}$  表示第  $i$  个交巡警平台封锁第  $j$  个交通要道时锁定的出警线路路程。

约束条件即为

$$\sum_{i=1}^{20} a_{ij} = 1, \quad j = 1, 2, \dots, 13 \quad \text{表示一个路口必须由一个交巡警平台来实施封锁;}$$

$$\sum_{j=1}^{13} a_{ij} \leq 1, \quad i = 1, 2, \dots, 20 \quad \text{表示一个交巡警平台最多只封锁一个路口。}$$

利用Lingo软件编程  $c = 8015 \text{ min}$ , 即交巡警服务平台最快可在8015分钟内对13条交通要道进行全封锁, 最佳调度方案如表3所示。在本方案中, 交巡警出警所走总路程为48188mm。

表3 最佳调度方案

调度方案	调度路径	调度方案	调度路径
2 → 38	2 → 40 → 39 → 38	4 → 62	4 → 62
5 → 48	5 → 47 → 48	7 → 29	7 → 30 → 29
8 → 30	8 → 33 → 32 → 7 → 30	9 → 16	9 → 35 → 36 → 16
10 → 22	10 → 26 → 11 → 22	11 → 24	11 → 25 → 24
12 → 12	12 → 12	13 → 23	13 → 23
14 → 21	14 → 21	15 → 28	15 → 28
16 → 14	16 → 14		



### 2.1.3 增加平台模型(工作量均衡模型)的建立与求解

将3分钟内在交巡警到达事发地作为增加交巡警平台的约束条件, 即  $t_{ij} \leq 30 \text{ min}$

交巡警一次出警的工作量用其出警途中耗费和 $T$ 和处理案发现场耗时间 $T'$ 之和来表述, 那么一个服务平台的工作量即为  $W_i = \sum (T_i + T'_i)p_i$ . 工作量均衡以总方差<sup>[5]</sup>表示, 得目标函数

$$\min S^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2.$$

综上可建立如下模型

$$\begin{cases} \min S^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 \\ W_i = \sum (T_i + T'_i)p_i, \\ t_{ij} \leq 30, \\ 2 \leq n \leq 5. \end{cases}$$

在现有交巡警服务平台设置情况中由2.1.1可知有6个路口不符合约束条件, 因此增加的平台必须存在, 这6个路口周围, 进而由约束条件寻找可能增设平台的位置。通过分析, 至少需要分别从{28, 29} {38, 39, 40} {48, 61} {87, 88, 89, 90, 91, 92}这4组路口集合中选择4个路口增设才能满足设置标准。根据已建立的模型, 运用Matlab软件优化结果, 得到满足要求的位置为28, 39, 48, 87这四个路口。此时其工作量方差最小, 为5.6183。在此基础上再增设一个平台来进一步优化分配管辖范围, 做法与寻找4个点的情况相似, 最后得到一个最优的增设平台位置为69。相对4个增设平台位置的情况, 再增设一个平台的工作量最小方差为4.2925, 优化效果不明显。另外, 考虑到警备资源、经济状况等实际问题, 所以无需增设5个交巡警服务平台。因此在28, 39, 48, 87这四个路口增设新的交巡警服务平台为较合理的方案。

由此得到增设平台后, 交巡警服务平台的管辖范围, 如表4所示。

表4. 增设平台后管辖范围分配方案

平台	管辖范围	平台	管辖范围	平台	管辖范围
1	1, 67, 68, 69, 71, 73 74, 75, 76, 78	2	2, 43, 44, 70, 72	3	3, 54, 55, 65, 66
4	4, 57, 60, 62, 63, 64	5	5, 49, 50, 51, 52, 53, 56, 58, 59	6	6
7	7, 30, 32	8	8, 33, 46	9	9, 31, 34, 35, 45
10	10	11	11, 26, 27	12	12, 25
13	13, 21, 22, 23, 24	14	14	15	15, 28, 29
16	16, 36, 37	17	17, 41, 42	18	18, 80, 81, 82, 83
19	19, 77, 79	20	20, 85, 86	28	28, 29
39	38, 39	48	47, 48, 61	87	84, 87, 88, 89, 90, 91, 92

## 2.2 问题二模型的建立与求解

### 2.2.1 现有平台设置合理性分析

为分析现有平台设置的合理性,用 Floyd 算法求得全市内所有路口到达最近服务平台的最短路程。发现在多达 138 个路口发生案件时 3 分钟内交巡警无法到达,这严重不符合处理案件的及时性这一要求。考虑以 3 分钟内有交巡警可以到达案发现场为约束条件,以设置平台数最少为目标函数,在现有平台设置的基础上增设服务平台(方案 1)或不考虑现有平台重新设置服务平台(方案 2)两种方案来求得需要设置的平台数。

同时为了让每个服务平台的工作量均衡,用 2.1.3 中工作量均衡模型对平台设置的具体位置及管辖范围求解。

### 2.2.2 需增加平台数模型建立

基于 2.2.1 分析,新增平台后,应满足所有路口发生案件时 3 分钟内有交巡警,以到达案发现场。

**模型一** 在不改变现有平台布局的情况下增设平台,使所有路口都能在 3 分钟内有巡警到达,并使各平台工作量尽量均衡。建立如下模型。

$$\min \sum_{i=1}^m p_i - n$$

$$\begin{cases} l_{ij}, x_{ij} \leq 30, & i = 1, 2, \dots, m, \quad j = 1, 2, \dots, m, & (a) \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, m, & (b) \\ x_{ij} \leq p_i, & i = 1, 2, \dots, m, \quad j = 1, 2, \dots, m, & (c) \\ x_{ij} = 0 \text{ 或 } 1, & i = 1, 2, \dots, m, & (d) \\ p_i = 0 \text{ 或 } 1, & i = 1, 2, \dots, m, & (e) \\ p_i = 1, & i = 1, 2, \dots, n, & (f) \end{cases}$$

其中  $n$  为已有平台数,考虑到警力资源的有限性,将增加平台数  $\sum_{i=1}^m p_i - n$  作为目标函数,这时

$$p_i = \begin{cases} 1, & \text{第 } i \text{ 个路口是平台,} \\ 0, & \text{第 } i \text{ 个路口不是平台} \end{cases}$$

$m$  为总路口数,  $l_{ij}$  第  $i$  个平台到第  $j$  个路口的最短路程

$$x_{ij} = \begin{cases} 1, & \text{第 } i \text{ 个平台管辖第 } j \text{ 个路口,} \\ 0, & \text{第 } i \text{ 个平台不管辖第 } j \text{ 个路口} \end{cases}$$

(a) 式表示每个平台管辖的所有点能够 3 分钟赶到, (b) 式表示每个路口只能由一个平台管辖, (c) 式表示只有在第  $i$  个路口为服务平台时才能将第  $j$  个路口让其管辖, (f) 式表示前  $n$  个路口为已有平台。

此模型确定了增加平台的个数,而具体增设位置及管辖范围应考虑到要尽量使平台的工作量均衡,用 5.1.3 中的工作量均衡模型求解。

用 Lingo 编程求得增设平台情况,如表 5 所示。

表5: 解决方案一

区域	新增平台个数	新增平台位置
A	4	28, 39, 48, 87
B	2	104, 149
C	15	184, 199, 201, 204, 205, 216, 238, 240, 252, 257, 259, 263, 288, 304, 318
D	8	329, 332, 333, 338, 344, 362, 389, 371
E	14	387, 389, 390, 392, 393, 408, 419, 420, 421, 446, 454, 458, 472, 474
F	13	491, 509, 513, 517, 539, 541, 558, 560, 569, 574, 577, 578, 582

模型二 不考虑已有平台, 对所有路口进行重新布局, 使所有路口都能在3分钟内在交警巡逻警到达, 并使各平台工作量尽量均衡. 类似模型一, 可建立如下模型:

$$\min \sum_{i=1}^m p_i$$
$$\begin{cases} l_{ij}x_{ij} \leq 30, & i=1,2,\cdots,m, \quad j=1,2,\cdots,m, \\ \sum_{i=1}^m x_{ij} = 1, & j=1,2,\cdots,m, \\ x_{ij} \leq p_i, & i=1,2,\cdots,m, \quad j=1,2,\cdots,m, \\ x_{ij} = 0 \text{ 或 } 1, & i=1,2,\cdots,m, \\ p_i = 0 \text{ 或 } 1, & i=1,2,\cdots,m. \end{cases}$$

用Lingo编程求得增设平台情况, 如表6所示.

表6: 解决方案二

区域	新增平台个数	新增平台位置
A	15	3, 10, 11, 14, 15, 22, 25, 29, 33, 40, 43, 48, 58, 66, 90
B	5	104, 114, 135, 149, 163
C	27	166, 175, 177, 179, 184, 185, 193, 199, 201, 204, 205, 214, 222, 231, 238, 240, 250, 254, 256, 259, 263, 266, 274, 285, 303, 310, 319
D	12	327, 329, 332, 333, 338, 344, 350, 361, 362, 363, 369, 371
E	23	373, 378, 387, 389, 390, 393, 397, 400, 408, 416, 419, 420, 421, 423, 424, 440, 444, 446, 457, 459, 470, 472, 473
F	19	479, 485, 488, 489, 499, 509, 511, 515, 532, 539, 541, 559, 562, 565, 570, 573, 574, 575, 582

比较模型一与模型二, 发现模型二只需设101个平台, 而模型一需增加56个, 即共设157个平台. 从警力资源的有限性考虑, 选择模型二结果作为调整方案.

2.2.3 最佳围堵问题模型的建立与求解

此问题可类似于2.1.2.4的警力合理调度问题，不同的是需要调度全市警力围堵封锁，而完成围堵封锁的最短时间必须小于嫌疑人逃跑时间减去3分钟，即

{ min c = [ max\_{1 ≤ j ≤ n} ( ∑\_{i=1}^{80} e\_{ij} ) ] } < t - 3

因此可建立模型如下：

目标函数为

min t,

约束条件为

{ min c = [ max\_{1 ≤ j ≤ n} ( ∑\_{i=1}^{80} e\_{ij} ) ] ≤ t - 3, vt < T\_{vj}, e\_{ij} = a\_{ij}t\_{ij}, a\_{ij} = 0 或 1, ∑\_{i=1}^{80} a\_{ij} = 1, ∑\_{j=1}^n a\_{ij} ≤ 1.

利用Lingo编程求解可得最佳围堵方案，如表7所示。在此方案中，接到报案后7.36min就形成包围圈，满足快速搜捕要求，为最佳围堵方案。

表7：围堵具体方案

序号	围堵路径	序号	围堵路径	序号	围堵路径
1	13, 22, 21, 14	3	17, 17	5	12, 37, 26
4	15, 28, 29	5	2, 43, 42, 27, 41	6	3, 44, 2, 43
7	5, 50, 51, 59, 60, 62	8	1, 69, 68	9	10, 34, 35, 45, 44, 2, 70
10	19, 77, 76	11	4, 62, 190, 189, 168	12	73, 232, 231, 216, 215
13	171, 228, 218	14	172, 227	15	169, 254, 253, 240
16	167, 249, 248	17	174, 224, 223, 225, 170, 273	18	320, 349, 371
19	481, 491, 486, 487, 489, 482	20	477, 498, 495, 493, 492, 491, 486, 487	21	476, 545, 546, 547, 532, 548, 549
22	478, 554, 555, 557, 558	23	475, 557, 564, 563, 562		

### 3 模型的评价

- 1) 建立的模型简洁明了, 使用了数学工具 Matlab、Lingo 等软件, 提高了工作效率
  - 2) 修改模型参数可以满足各种同类需求, 具有很强通用性
- 当然模型还存在许多需要改进的地方。例如可考虑车速变化等问题。

### 参考文献

- 1 吴新宗. 运筹学: 最优化方法[M]. 北京: 机械工业出版社, 2003.  
Wu X Z. Operations Research and the Optimization Method[M]. Beijing: Mechanical Industry Press, 2003.
- 2 宋叶忠, 罗东明. Matlab 数值分析[M]. 北京: 机械工业出版社, 2009.  
Song Y Z, Luo D Y. Numerical Analysis and Application of Matlab[M]. Beijing: Mechanical Industry Press, 2009.
- 3 林世松, 周敏, 陈建雷. Matlab7.0 应用案例[M]. 北京: 机械工业出版社, 2006.  
Lin X S, Zhou Q, Chen D X. Matlab7.0 Application[M]. Beijing: Mechanical Industry Press, 2006.
- 4 江启源, 谢金平, 叶茂. 数学模型[M]. 北京: 高等教育出版社, 2004.  
Jiang Q Y, Xie J X, Ye J. Mathematical Model[M]. Beijing: Higher Education Press, 2004.
- 5 韩志超. 数学建模与优化[M]. 北京: 高等教育出版社, 2005.  
Han Z C. Mathematical Modeling and Application[M]. Beijing: Higher Education Press, 2005.

## Setting and Scheduling of Traffic and Patrol Police Service Platforms

HU Li-zhong, ZHANG Jin-lei, ZHU Jin-bao

Advisor: LI An-ping, AN Qin-li, LIANG Fang-chi

Missile Institute, Air Force Engineering University, Xi'an 713800

**Abstract:** It is an important issue to allocate and schedule the traffic and patrol police service platforms (TPPSPs) in accordance with the practical police force and condition. By employing the Floyd algorithm, the distances between the nodes and TPPSPs are obtained and the jurisdiction region for each TPPSP in the city center District A is allocated. Based on the objective function of minimizing the blockade time and the shortcut, the blockade scheme for the TPPSPs of District A is scheduled. The supplementary TPPSPs number is accessed on condition that the police can arrive at the locale within 3 minutes with the minimum workman variance of the TPPSPs of District A and the minimum number of supplemental TPPSPs. Through investigating the reasonableness of the TPPSP setting for the entire city, the 118 access arteries cannot arrive at the locale within 3 minutes. The models of added and reallocated TPPSPs are adjusted on condition that the police can arrive at the locale within 3 minutes. An optimal containment strategy is established according to the objective function of minimum blockade time, region area and region population.

**Keywords:** 0-1 program; Floyd algorithm; short path problem; balance degree



### 3 模型的建立与求解

#### 3.1 问题 A 区交巡警平台的设置与调度分析

##### 3.1.1 A 区交巡警平台的管辖范围分配

当出现突发事件时,为使交巡警力尽量在 3 分钟内,警车的时速为 60 km/h 到达事发地点,需要各路路口距离最近的交巡警平台来管辖.首先,根据 Floyd 算法,计算 11 个平台与各路路口路径距离最小值.再结合图 1 中的流程图,利用 Matlab 编程<sup>[4]</sup>,寻找距离各路路口最近的平台并计算其距离.

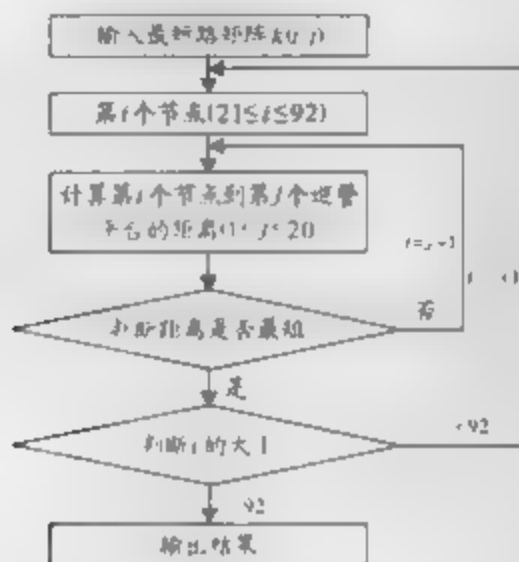


图 1. A 区寻找距离各路路口最近的交巡警平台的流程图

由此可得各平台的管辖范围,见表 1.

表 1: 各平台的管辖范围

交巡警平台	节点	交巡警平台	节点
A1	1 67 68 69 71 73 74 75 76 78	A11	11 26 27
A2	2 39 40 43 44 70 72	A12	12 25
A3	3 54 55 65 66	A13	13 21 22 23 24
A4	4 57 60 62 63 64	A14	14
A5	5 49 50 51 52 53 56 58 59	A15	15 28 29
A6	6	A16	16 36 37 38
A7	7 30 32 47 48 61	A17	17 41 42
A8	8 33 45	A18	18 80 81 82 83
A9	9 31 34 35 45	A19	19 77 79
A10	10	A20	20 84 85 86 87 88 89 90 91 92

### 3.1.2 A区13条交通要道的快速封锁调度方案

发生重大突发事件后, 调度20个交巡警服务平台的警力资源, 对进出该区的13条交通要道实现快速全封锁。实际中, 一个平台的警力最多封锁一个路口, 因此这属于指派问题。

对于各平台, 只有调度和不调度两种情况, 因此, 可用0-1规划的思想建立模型<sup>[3]</sup>。设 $x_{kj}$ 为第 $k$ 个出口被第 $j$ 个平台的警力封锁的情况, 则有

$$x_{kj} = \begin{cases} 0, & k=1, 2, \dots, l, \quad j=1, 2, \dots, n, \\ 1, & \end{cases} \quad (1)$$

其中 $x_{kj}=0$ 表示第 $k$ 个出口不被第 $j$ 个平台的警力封锁,  $x_{kj}=1$ 表示第 $k$ 个出口被第 $j$ 个平台的警力封锁。  $n$ 表示研究范围内交巡警平台的个数,  $l$ 表示研究范围内出口个数。

为了在最短时间内实现全封锁, 而全封锁的时间是由封锁最后一个路口所用的时间决定的。因此, 目标函数就是求最远调度距离的最小值, 表述为

$$\min Z = \max \left( \sum_{j=1}^n S_{kj} x_{kj} \right), \quad (2)$$

其中 $Z$ 表示所有调度中的最远距离,  $S_{kj}$ 表示第 $j$ 个平台到第 $k$ 个出口的距离, 该距离由Floyd算法得出。

约束条件为:

1) 平台不一定被调去封锁出口, 即

$$\sum_{k=1}^l x_{kj} \leq 1 \quad j=1, 2, \dots, n \quad (3)$$

2) 出口只被一个平台封锁, 则有

$$\sum_{j=1}^n x_{kj} = 1 \quad k=1, 2, \dots, l \quad (4)$$

综上, 最快实现全封锁的模型为<sup>[4]</sup>:

$$\begin{aligned} \min Z &= \max \left( \sum_{j=1}^n S_{kj} x_{kj} \right), \\ \text{s.t.} \quad &\begin{cases} \sum_{k=1}^l x_{kj} \leq 1, & j=1, 2, \dots, n, \\ \sum_{j=1}^n x_{kj} = 1, & k=1, 2, \dots, l, \\ x_{kj} \in \{0, 1\} \end{cases} \end{aligned} \quad (5)$$

根据模型(5), 利用Matlab编程和题目中给出的相关数据, 最后可以得到若干组解。这些解的共同特点是最远的调度距离都是平台7至出口9的距离, 为80.15百米。所以完成A区全封锁的最短时间即由此决定, 需要8分钟。

为了得到最优的调度方案, 在此基础上, 增加“最远的调度距离是平台7至出口9的距离”这一约束条件, 以总调度距离最短为目标函数, 对除平台7和出口9以外的出口和交巡警平台



进一步作 0-1 规划, 模型为

$$\begin{aligned} \min Y &= \sum_{k=1}^{l-1} \sum_{j=1}^{n-1} S_{kj} x_{kj} \\ \text{s.t.} &\begin{cases} \max(\sum_{j=1}^n S_{kj} x_{kj}) < S_{97}, \\ \sum_{k=1}^l x_{kj} < 1, \quad j=1, 2, \dots, n-1 \\ \sum_{j=1}^n x_{kj} = 1 \quad k=1, 2, \dots, l-1 \\ x_{kj} \in \{0, 1\}. \end{cases} \end{aligned} \quad (6)$$

其中  $Y$  表示除平台 7 到出口 9 距离以外的总调度距离,  $(n-1)$  表示除平台 7 以外的平台总数,  $(l-1)$  表示除出口 9 以外的出口总数.

利用 Lingo 软件编程求解<sup>[5]</sup>, 最终的调度方案见表 2.

表 2 最快实现完全封锁且总距离相对最短的调度方案

出口	1	2	3	4	5	6	7	8	9	10	11	12	13
平台	A12	A16	A8	A14	A10	A13	A11	A15	A7	A9	A2	A5	A4
距离/千米	0.00	07.42	26.92	32.65	77.06	5.00	38.05	47.52	80.15	34.92	39.82	24.76	3.50

由表 2 可知, 最快实现完全封锁的时间为 8 分钟, 调度的总距离为 477.79 千米.

### 3.1.3 增加平台后管辖范围的分配方案

由于各平台管辖范围内的路口数差异很大, 以及各路口的案发率不同, 造成现有交巡警服务平台的工作量不平衡, 部分地方的出警时间过长. 首先, 在不增加交巡警服务平台的前提下重新分配管辖范围.

对于路口  $i$  有被平台  $j$  管辖和不被平台  $j$  管辖两种情况, 因此, 可设 0-1 变量  $x_{ij}$  令

$$x_{ij} = \begin{cases} 0, & i=1, 2, \dots, m, \quad j=1, 2, \dots, n, \\ 1, & \end{cases} \quad (7)$$

其中  $x_{ij} = 0$  表示第  $i$  个路口不被第  $j$  个平台管辖,  $x_{ij} = 1$  表示第  $i$  个路口被第  $j$  个平台管辖.  $n$  表示研究范围内交巡警平台的个数,  $m$  表示研究范围内路口的个数.

管辖区域的划分原则为使各平台的工作量尽量均衡, 出警时间尽量缩短.

**目标函数 1** 以各平台工作量的变异系数来表征工作量的均衡性, 其值越小表示工作量越均衡. 首先, 各交巡警平台的工作量可表示为该平台管辖范围内所有路口案发率的总和, 记

$$W_j = \sum_{i=1}^m C_i x_{ij}, \quad (8)$$

其中  $W_j$  指交巡警平台  $j$  的工作量,  $C_i$  表示路口  $i$  的日案发率.

所有平台工作量的平均值为

$$\bar{W} = \frac{1}{n} \sum_{j=1}^n W_j, \quad (9)$$

则各平台工作量的变异系数最小表示为

$$\min f_1 = \frac{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (W_j - \bar{W})^2}}{\bar{W}} \quad (10)$$

**目标函数2** 所有平台的最长出警时间相比, 其中的最大值最小表示最长出警时间最短。交巡警的出警时间是指从接警到达事发地路口的时间, 则第  $j$  个平台的最长出警时间为

$$T_j = \max \left( \frac{S_{ij} x_{ij}}{V} \right), \quad (11)$$

其中  $S_{ij}$  表示第  $j$  个平台到达第  $i$  个路口, 的最短距离。则最长出警时间最短表示为

$$\min f_2 = \max(T_j). \quad (12)$$

约束条件如下:

1) 平台不闲的约束 为使每个平台不至于无管辖范围, 可约束为已至少管辖自己所在的路口, 当  $i=j$  时, 即

$$x_{ij} = 1, \quad i = j. \quad (13)$$

2) 每个路口都被平台管辖的约束 当  $i \neq j$  时, 第  $i$  个路口必定被  $1 \sim n$  中的唯一一个平台管辖, 即

$$\sum_{j=1}^n x_{ij} = 1, \quad i \neq j. \quad (14)$$

3) 出警时间不超过 3 min, 即

$$\frac{S_{ij}}{V} \leq 3. \quad (15)$$

综上, 考虑平台的工作量均衡及出警时间合理的模型可表示为

$$\begin{aligned} \min f &= \frac{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (W_j - \bar{W})^2}}{\bar{W}} \quad \min f_2 = \max(T_j) \\ \text{s.t.} &\begin{cases} x_{ij} = 1, & i = j, \\ \sum_{j=1}^n x_{ij} = 1, & i \neq j, \\ \frac{S_{ij}}{V} \leq 3, \\ x_{ij} \in \{0, 1\} \end{cases} \end{aligned} \quad (16)$$

在不增加交巡警平台的前提下, 根据模型 (16), 利用 Matlab 软件进行求解, 结果见表 3。据此可知在此前提下, 最长出警时间为 5.70 min, 出现在平台 15 前往路口 29 处理突发事件时, 工作量的变异系数为 0.1830。

为缩短最长出警时间, 将新增平台设置在出警时间最长的路口处, 同理可求得增加平台 1 到 5 个时工作量变异系数及最长出警时间的变化, 见表 4。

**结论** 增加 4 个交巡警平台时, 分别位于路口 28(29)、61、39 和 91, 此时最长出警时间已达到 3 min 以内的要求, 为 2.71 min, 工作量的变异系数较小, 为 0.2004。增加 5 个交巡警平台

时, 分别位于路128(29)、61、39、91和67, 1件量的变异系数减小, 为0.1526, 最长出警时间仍为2.71 min。因此, 增加4或5个交巡警平台都满足题目要求。考虑到建设成本时, 可只增加4个交巡警平台, 若更重视1件量的均衡性, 需增加5个交巡警平台。

表3 最长出警时间最短且工作量均衡时各平台的管辖范围

交巡警平台	管辖的路口	日工作量(案件数)
A1	1 71 73 74 75 68	6.5
A2	2 43 44 70 69	6.9
A3	3 54 55 55 66 67	6.4
A4	4 57 60 62 63 64	6.6
A5	5 48 52 53 56 58	6.9
A6	6 50 59 47 51 48	6.9
A7	7 30 61	5.1
A8	8 33 46 32	6.5
A9	9 31 35 45	6.5
A10	10 34	3.7
A11	11 26 27	5.6
A12	12 25 24	5.1
A13	13 22 23	5.0
A14	14 21	4.9
A15	15 28 29	4.8
A16	16 36 37 38 39	6.4
A17	17 41 42 40 72	6.8
A18	18 81 82 83 84 80 86	8.4
A19	19 76 77 78 79 80	6.1
A20	20 87 88 89 91 92 85	8.4

表4 增加平台后工作量的变异系数和最长出警时间

新增平台个数	新增平台位置	工作量的标准差	工作量的均值	变异系数	最长出警时间
0	无	1.14	6.23	0.1830	5.70
1	28或29	1.34	5.93	0.2260	4.19
2	61	1.51	5.66	0.2668	3.82
3	39	1.35	5.41	0.2495	3.68
4	91	1.04	5.19	0.2004	2.71
5	67	0.76	4.98	0.1526	2.71

### 3.2 问题二:全市交巡警平台的设置与调度

#### 3.2.1 全市现有交巡警平台设置的合理性分析及调整方案

首先,在不增加交巡警平台时,根据问题一中的模型(16)对主城六区分别进行分析,结果见表5

表5: 全市现有交巡警平台的相关数据

主城六区	路口数	平台数	平均工作量	变异系数	最长出警时间
A区	92	20	6.23	0.1830	5.70
B区	73	8	6.38	0.1743	4.47
C区	154	17	8.85	0.1725	6.86
D区	52	9	5.24	0.2070	16.06
E区	103	15	5.79	0.1979	19.10
F区	108	11	5.60	0.1883	8.48

根据表5中的数据,我们对全市(A, B, C, D, E, F)现有的交巡警服务平台设置方案的合理性进行分析。结果发现:

1, 主城各区工作量的变异系数都较小,即各交巡警平台的工作量均衡,比较合理;

2, 主城各区的最长出警时间都较大,尤其是D区和E区,远远超过了规定的出警时间3min,不合理,因此各区的交巡警服务平台都有待调整。造成这一结果的原因为部分路口与最近平台间的距离超过3km。

针对全市各区交巡警平台的出警时间过长这一问题,我们在不改变原有交巡警平台位置的基础上,增加必要的平台,使最长出警时间小于3min

利用问题一A区增加平台的方法,可分别求出B、C、D、E、F各区新增平台的个数及位置,并进一步求解得增加平台后各区交巡警平台的最长出警时间,结果见表6

表6: 增加平台位置及增加前后各区最长出警时间的比较(单位:分钟)

主城六区	不增加平台时	新增平台(路口编号)	增加平台后
A区	5.70	26 30 61 92	2.71
B区	4.47	123 152	2.91
C区	6.86	166 167 168 169 170 171 174 175 176 177 178 179 180 183 199 201 203 206	2.99
D区	16.06	320 322 324 325 326 328 329	2.91
E区	19.10	372 373 374 376 378 379 380 381 382 383 384 385 386 387 388 390 393	2.97
F区	8.48	475 477 478 479 480 482 483 484 485 486 490 505	2.95

由表6可知,新增平台后,主城六区的最长出警时间与不增加平台时相比,明显缩短,全部控制在3min之内,因此更合理

### 3.2.2 最佳围堵方案

#### 3.2.2.1 模型的建立

以最快抓捕嫌疑犯为目标, 表述为

$$\min U = \max \left( \frac{L(i, 32)}{V} \right), \quad (17)$$

其中  $L(i, 32)$  表示从事发地—路口32到落网地点—路口 $i$ 的距离,  $U$ 表示嫌疑犯从逃跑到落网的时间, 根据假设8, 嫌疑犯的逃跑速度为 $V$

约束条件为,

1) 由于部分路口 $i$ 的封堵对抓捕嫌疑犯无意义, 因此, 不一定每个路口都被封堵, 即

$$\sum_{j=1}^n x_{ij} \leq 1, \quad i = 1, 2, \dots, m, \quad (18)$$

$$x_{ij} = \begin{cases} 0, & i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n, \\ 1, & \end{cases} \quad (19)$$

其中  $x_{ij} = 0$  表示第 $i$ 个路口没有被第 $j$ 个平台封堵,  $x_{ij} = 1$  表示第 $i$ 个路口被第 $j$ 个平台封堵,  $n$ 表示研究范围内交巡警中 $\tau_i$ 的个数,  $m$ 表示研究范围内路口 $i$ 的个数

2) 不一定每个平台都出警, 即

$$\sum_{i=1}^m x_{ij} \leq 1, \quad j = 1, 2, \dots, n. \quad (20)$$

3) 嫌疑犯到达路口 $i$ 的时间减3min比交巡警从平台 $j$ 到路口 $i$ 的时间长, 表示路口 $i$ 能被 $\tau_j$ 的被平台 $j$ 封堵, 即

$$\frac{L(i, j)}{V} \leq \frac{L(i, 32)}{V} - 3, \quad (21)$$

其中  $L(i, j)$  表示平台 $j$ 到路口 $i$ 的距离,

4) 为描述嫌疑犯被封锁在一定的区域内, 定义路口 $i$ 被封锁为  $L'(Q_1) = \inf$ , 即任意路口 $i$ 与路口 $i$ 不相通, 距离表示为  $\inf$ . 若  $Q_1$  表示不被封堵,  $x$  域以外所有路口 $i$ 的集合,  $Q_2$  表示被封锁区域内所有路口的集合, 则

$$L'(Q_1, Q_2) = \inf. \quad (22)$$

综上, 模型为

$$\begin{aligned} \min U &= \max \left( \frac{L(i, 32)}{V} \right), \\ \text{s.t.} \quad &\begin{cases} \sum_{j=1}^n x_{ij} \leq 1, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m x_{ij} \leq 1, \quad j = 1, 2, \dots, n, \\ \frac{L(i, j)}{V} < \frac{L(i, 32)}{V} - 3, \\ L'(Q_1, Q_2) = \inf, \\ x_{ij} \in \{0, 1\} \end{cases} \end{aligned} \quad (23)$$

### 3.2.2.2 模型的求解

将相关数据代入模型(23), 可得到最终的封堵方案, 见表7

表7 全市各区的围堵方案

调度 平台	封锁 路口	逃跑 时间	封锁 时间	调度 平台	封锁 路口	逃跑 时间	封锁 时间	调度 平台	封锁 路口	逃跑 时间	封锁 时间
A1	63	8.63	3.50	A16	16	3.30	0.00	C4	240	10.15	6.94
A2	3	8.48	2.11	A17	40	7.95	2.69	C6	245	5.61	3.57
A3	55	5.21	1.27	A18	41	10.50	5.54	C7	231	6.98	2.78
A4	4	8.80	0.00	A19	62	9.13	5.03	C8	246	6.54	3.08
A10	10	6.19	0.00	C2	248	20.52	3.67	D1	371	15.89	7.36
A15	29	9.16	5.70	C3	168	12.59	0.00	F1	561	8.79	4.35

完成以上所有路口的封锁后, 即可成功围堵嫌疑犯。完成全区域封锁的时间为7.36分钟, 嫌疑犯最迟会在20.52分钟被成功拦截。

### 参考文献:

- [1] 叶奇明, 于世光. Floyd算法的数学模型研究[J]. 海南大学学报(自然科学版), 2008, 26(1): 47-50.  
Ye Q M, Shi S G. The research of Floyd algorithm detour model[J]. Journal of Hainan University (Natural Science), 2008, 26(1): 47-50.
- [2] 何勇. Matlab数学实验与建模[M]. 北京: 清华大学出版社, 2010.  
He Y. Matlab Mathematical Experiments and Modeling[M]. Beijing: Tsinghua University Press, 2010.
- [3] 朱炎, 李利, 周桂梅. 智能交通系统导论[M]. 北京: 中国人民大学出版社, 2007.  
Zhu Y, Wang J L, Zhou T M. Introduction to Intelligent Transport System[M]. Beijing: Chinese People's Public Security University Press, 2007.
- [4] 姜启源, 谢金荣, 叶俊. 数学模型(第3版)[M]. 北京: 高等教育出版社, 2003.  
Jiang Q Y, Xie J X, Ye J. Mathematical Modeling (3rd Edition)[M]. Beijing: Higher Education Press, 2003.
- [5] 谢金荣. 优化建模——Lingo软件[M]. 北京: 清华大学出版社, 2011.  
Xie J X. Optimization Modeling and Lingo Lingo Software[M]. Beijing: Tsinghua University Press, 2011.
- [6] 张楠, 王坤. 流线网络优化的变分不等式模型与算法[J]. 西南交通大学学报, 2011, 46(3): 481-487.  
Zhang J, Wang K. Variational inequality model and algorithm for stream line network optimization[J]. Journal of Southwest Jiaotong University, 2011, 46(3): 481-487.

## The Setting and Scheduling Model of Traffic and Patrol Police Platform Based on 0-1 Programming

TANG Di<sup>1</sup>, DONG Xiao-xiao<sup>1</sup>, WEI Xin<sup>1</sup>

Advisor: LUO Wan-chun<sup>2</sup>

1- The 9th Team of Guard Brigade Third Military Medical University, Chongqing 400038;

2- Department of Mathematics, Third Military Medical University, Chongqing 400038

**Abstract:** This paper aims at resolving the problem of the setting and scheduling of the traffic and patrol police platform. First, the shortest path and distance between the intersection and platform is calculated based on the Floyd algorithm, and each intersection is set to be dominated by the nearest platform. In emergency case, the model is established on the basis of 0-1 programming to attain the fastest and the shortest scheduling scheme. By compromising the workload of each platform and the time wasted on the way to the intersection into consideration, we establish a partition model for the territorial jurisdiction and achieve the scheme of setting new platforms. Finally, in order to catch the criminal fled from Point P as soon as possible, we construct the optimization model by setting the shortest longest fleeing time to be shortest as the objective and that the police can successfully close the intersections and the criminal can be completely blocked up as the constraint conditions. The best plan for blockade can then be drawn by solving the model using 0-1 programming. The final result is as follows: the police from 8 platforms should be scheduled to close 18 intersections and the criminal is to be caught within 20.52 minutes.

**Keywords:** shortest path, 0-1 integer programming, the police platform





分析表1数据,发现最短路径算法下的工作量极不均衡,因此提出优化方案——最短路径工作量方案

表1: 最短路径下A的辖区分配

平台号	辖区内的节点号	工作量	平台号	辖区内的节点号	工作量
1	1 67 68 69 71 73 74 75 76 78	10.3	11	11 26 27	4.6
2	2 39 40 43 44 70 72	9.7	12	12 25	4
3	3 54 55 65 66	5.6	13	13 21 22 23 24	8.5
4	4 57 60 62 63 64	6.6	14	14	2.5
5	5 49 50 51 52 53 56 58 59	9.7	15	15 28 29	4.4
6	6	2.5	16	16 36 37 38	5
7	7 31 32 47 48 61	9.6	17	17 41 42	5.3
8	8 33 46	5	18	18 80 81 82 83	6.1
9	9 31 34 35 45	8.2	19	19 77 79	3.4
10	10	1.6	20	20 64 85 86 87 88 89 90 91 92	11.5

### 3.1.2 考虑工作量均衡下的最短路径辖区划分方案的进一步优化

改进方案的调整本身就是一个协调,因此在筛选工作量较大的平台时,对筛选标准进行简化,以平台管辖 $\times$ 内的节点个数为标准,筛选出工作量过大的几个平台

对1, 6个平台,选与距离平台最近的几个节点,将这几个节点与周边其他平台进行归配,归配的条件是:满足“3分钟阈值”以及目标平台的节点数为饱和。如果有满足归配条件的平台,那么选择距离最近的平台,将管辖权转移;否则,节点仍归属原平台

利用新方案中的算法,对各平台内的管辖权进行调整,可以得到调整后的辖区分配表,如表2所示

表2: 调整后的辖区分配

平台号	辖区内的节点号	工作量	平台号	辖区内的节点号	工作量
1	1 69 71 74	5	11	11 26 27	4.6
2	2 40 43 44 72 70 68	9.2	12	12 25	4
3	3 54 55 65 66 67	6.4	13	13 22 23 24	7
4	4 57 60 62 63 64	6.6	14	14 21	3.9
5	49 50 51 52 53 56	7.7	15	15 28 29 31	6.4
6	6 47 58 59	6.1	16	16 36 37 38 39	6.4
7	7 30 48 61	6.1	17	17 41 42 92	6.1
8	8 33 46 32	6.5	18	18 81 82 83 88 73 90	8
9	9 34 35 45	6.6	19	19 77 79 80 76 78 75	7.9
10	10	1.6	20	20 85 86 87 89 91	7.9

### 3.1.3 工作量均衡前后的对比分析

均衡前后各平台工作量对比情况,如图1所示。调整前,工作量差

$$\sigma^2 = \sum_{i=1}^N (a_i - \mu)^2 = 8.009875,$$

均衡后工作量方差

$$\sigma^2 = \sum_{i=1}^{20} (a_i - \mu)^2 = 2.846475.$$

后者在工作量均衡度方面明显比前者更优。在最终方案1得到管辖区域划分的最终结果,见表4,如图2所示。

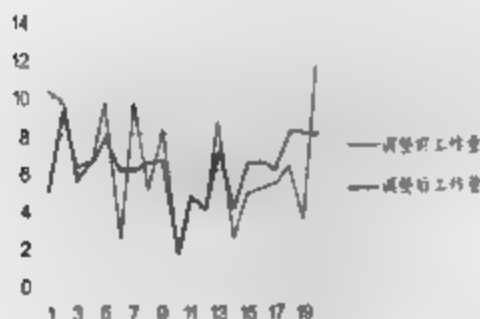


图1 调整前后各平台工作量对比

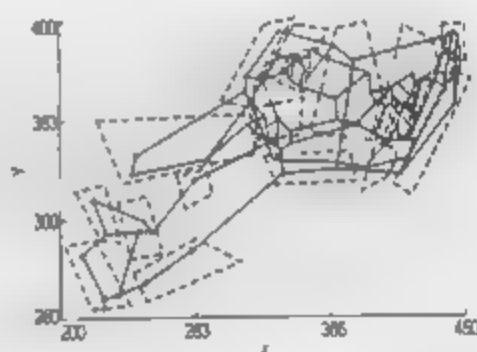


图2 最终辖区划分图

### 3.2 封锁方案的求解

以“最小化最大的最短距离”为标准,利用分支定界法使用最优性剪枝的搜索算法求得最优封锁路线。其中“最小化最大的最短距离”是指在所有封锁指派方案中,13个需要封锁的路口对应平台需要的报警时间的最大值在所有方案中最小。

最直接的方案是枚举法。将每种封锁方案搜索出来,筛选出最优方案。对于获得的多组最优方案的选择,采用多级目标<sup>[8]</sup>:

I级目标—最长封锁时间最短:  $\min \max D_{i, \text{map}_i}$ ;

II级目标—封锁平均耗时:  $\min \sum D_{i, \text{map}_i}$ ;

这里  $\text{map}_i$  为某个指派方案中,第  $i$  个要封锁的路口对应分配的交巡警平台的编号。

若枚举全部状态,则I级目标令枚举可知有  $13^6$  种方案,求每种方案中的距离采用Floyd算法求出的最短距离。

进一步优化可构造一种更加简洁的算法。分支定界法算法。基于分支的一些性质,可以求查找满足条件的临界点。查找的上界为10,查找的下界为给定区域中最远两点的距离,不妨令其为原图对角线的长度  $r = \sqrt{500^2 + 600^2} \approx 782$ 。显然,对于  $r$ ,必定可以出现全封锁。但仅靠上述算法不能求出具体的分配方案,接下来使用最优性剪枝搜索方法,可以把所有符合条件的方案找出来,从中选择满足II级目标的最优解,得到最终封锁模型示意图,如图3所示。计算得最长封锁时间为8.015分钟,为由7号平台处的警力去封锁29号节点这一段。

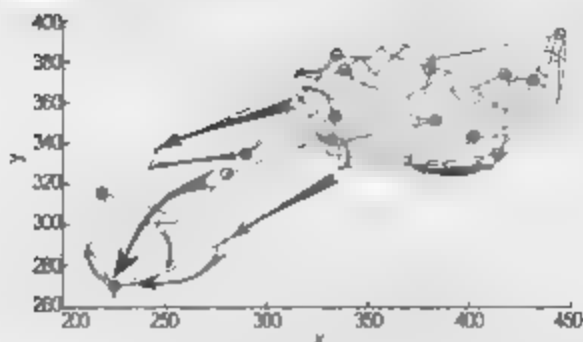


图3 巡逻模型示意图

### 3.3 新增交巡警平台位置确定

对A×路线车观察站发现, 东北区域的节点较多, 路段长度较短, 因此新增加交巡警平台可以分担工作量; 而西南区域的节点较少, 路段长度较长, 因此新增加交巡警平台可以解决民警巡逻长问题。故从节省人力资源的角度考虑, 在能较好的解决工作量与民警的目的的前提下, 选择新增平台数较少的更优。根据上述分析, 现建立新增交巡警平台位置模型。

#### 3.3.1 新增交巡警平台位置模型的建立

新增的交巡警平台是为解决两个问题: 第一, 工作量不均衡; 第二, 某些地区平台民警时间过长。

民警巡逻的衡量标准是: 是否在3分钟内到达距离自己最近的平台节点(都), 可以认为是“迅速”。将这样的路段选中, 在路段上点处插入节点, 交巡警平台可以设置在该点处, 可以起到缩短民警时间的作用。

假设现有交巡警平台数量为 $M$ , 位置确定, 路段节点数量为 $N$ 。

设各路段长度为 $l_i$ , 对于 $l_i > 600\text{m}$ 的路段, 在路段上点处插入节点。设新插入节点处的案发率为 $D$ , 这样插入的总节点数为 $n(n = 1, 2, \dots)$ 。新插入的节点与原来节点的地权元相等。总节点数为 $N + n$ 。设新增交巡警平台的数量为 $m$ 。

现在问题变为: 如何从 $N + n$ 个节点中选出 $m$ 个设置平台, 并对现有 $M + m$ 个平台的管辖区域适当划分, 使得警力分配资源更优。

定义工作量方差为 $\sigma^2$ , 最短的最长巡逻时间为 $z$ , 那么问题可以描述为 $\min\{\sigma^2, z\}$ 。

抽象的变量为: 置 $i = 0, 1, 2, \dots, n$ , 其中

$$\sigma^2 = \sum \frac{\text{工作量} - \text{均值}}{M + m}^2 \quad z = \min \max \frac{D_{i,j}}{t}$$

$\sigma^2$ : 为工作量方差和最短最长路这两个指标相关性的一个评价函数。这是一个双目标的优化问题, 两个无冲突性的目标并行, 第一目标为 $\min \sigma^2$ , 第二目标为 $z < 3\text{min}$ 。

#### 3.3.2 新增交巡警平台数量及位置的确定

由于工作量方差和最长最短路这两个指标相关性不是很强, 因此对其分别进行考虑。

例中对新增平台数 $m$ 的约束为 $2 < m < 5$ , 因此对 $m = 0, 1, \dots, 5$ 的情况分别进行讨论, 其中 $m = 0$ 仅作为一个原始参考数据,  $m = 1$ 作为辅助参考数据。

在算出了工作量方差和最短的最长巡逻时间两方面数据后, 综合两个目标, 以求得目标函数的最优解。即为了解决工作量分配不均及某些地方民警时间过长问题。本文选择增加几个

平台位置分别是28、36、48、57、81。对现有的25个平台重新分管辖区,得到新辖区内的报警情况及工作量,如表3所示。新增平台的直观示意图,如图4所示。

表3 新增平台前后的各关键数据对比

$m$	报警时间超过3分钟的数目	最长报警距离/km	平均报警时间/min	工作量/方差
0	6	5.760525	1.122	3.680675
5	0	2.708330	0.814	1.73244

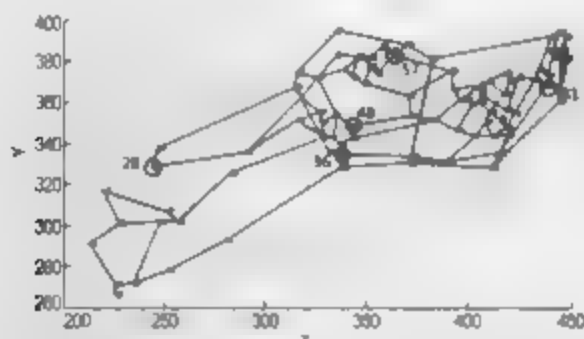


图4 新增平台示意图

### 3.4 全市交巡警平台设置方案的评价模型

设置交巡警服务平台的主要原则是:一旦发生有伤地起交通事故,发上重大案件时能及时对附近路段进行封锁,各个交巡警平台间工作量相对均衡。

全市的交巡警平台分布是A区分布的一个推广,所以用前文的模型对管辖区域做适当划分后,利用工作量、最长报警距离等指标,衡量平台设置是否合理。由于选取城市这个整体为对象,所需数据量大,因此我们下章结合建模讨论,最终将各数据整合,得到对全城的评价方案。

#### 3.4.1 分别对各城区建立评价模型

针对评价指标为:最长报警时间(无法实现3分钟覆盖的站点个数)、平均报警时间、工作量,先不考虑管辖区域的分配问题。此时1个目的为:作为这3个评价指标

在交巡警平台设置不合理又分为两种情况:一是平台数量不够,二是平台设置不合理。对前者,主要的解决方法是新建交巡警平台设置;对后者,主要的解决方法是调整现有的交巡警平台位置。当然,也有可能两种不合理的情况同时存在。但从节省人力资源的角度考虑,在能够调整交巡警平台位置的情况下,应尽量少新增平台数量。

1) 最短路径法下对时间的分析,见表4所示。

2) 对平均工作量的分析,见表5所示。

综合两种分析表格,以A区数据为参考指标,可以初步做出判断,如表6所示。

表 4 各地区出警情况统计

城 区	最短路算法下的时间量的分析					
	A	B	C	D	E	F
最长出警时间 (min)	57	44.7	68.6	160.6	191	84.8
无法 3 分钟到达节点个数	6	6	47	12	33	35
平均出警时间 min	11.22	15.27	23.35	25.06	25.38	24.3

表 5 各地区平均工作量统计

城 区	平均工作量					
	A	B	C	D	E	F
平均工作量	6.225	8.3	11.01	7.53	7.96	9.93

表 6 各地区解决方案

区域	主要问题	解决方案
B 区	情况与 A 区类似	增加 调整交巡警平台方案结合
C、F 区	交巡警平台工作量过大，在现有警力资源条件下 交巡警平台的负担过重	以增加新的交巡警平台为主
D、E 区	交巡警平台位置设置得不合理，因此产生部分节点无法被覆盖的情况	以调整现有交巡警平台位置为主

3.4.2 各城区评价及改进模型的建立

以 A × 为标准参照模板 对于各区的的评价标准可以参考 A × 地方案进行。

建立一个新增平台的评价模型，当新增平台数为  $r$  时，分别用  $x_i, \sigma_i^2, t_i$  表示工作期望量、工作量方差和平均出警时间。另外定义  $x_0, \sigma_0^2, t_0$  为初始值，即在新增平台数为 0 的情况下的工作期望量、工作量方差及平均出警时间。

$x_i$  衡量工作期望量， $\sigma_i^2$  衡量工作量分配的均衡度， $t_i$  衡量出警效率，将这三者加权值可以综合评价模型的优劣。因此我们定义一个评价函数

$$F=\lambda_1\frac{x_i}{x_0}+\lambda_2\frac{\sigma_i^2}{\sigma_0^2}+\lambda_3\frac{t_i}{t_0},$$

其中  $\lambda_1+\lambda_2+\lambda_3=1$  目标为  $\min F$   $\lambda_1, \lambda_2, \lambda_3$  的取值根据实际情况和具体要求确定(例如，若更注重工作量的均衡度，可取  $\lambda_1=\frac{1}{4}, \lambda_2=\frac{1}{2}, \lambda_3=\frac{1}{4}$ )

上述评价模型中没有涉及到最长出警时间的参数，因为分析表 4 的数据可知，最长出警时间与平均出警时间之间具有一定的相关性，而前者是特殊情况，后者则是基于整体的一个考虑，因此作为评价指标更有意义。

### 3.4.3 具体分析各个地区交巡警平台分布

先增加平台一个,利用前平台增加的模型,在未设置平台的节点总数 $N+n-1$ 的情况下,对新增的 $m$ 个平台的位置进行枚举,并且对每种方案计算出对应的平均区间 $F$ 值。选取 $F$ 值最小的方案。

以B区为例,取 $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$ ,计算出增加不同平台数时B警区的各项指标,并根据各项指标数据以及 $F$ 值的计算式,可画出 $F$ 值随平台增加个数变化的曲线图,如图5所示。

根据走势图,再结合节点数据,当增加平台数大于等于一个时,没有出警时间大于1分钟的节点,因此选择新增一个平台为最佳,这一个平台的位置是123、150、159。

同理,对C、D、E、F区类似处理,得到各城区应增加平台的节点统计表,见表7所示。绘制出新增平台后的平台设置图,如图6所示。

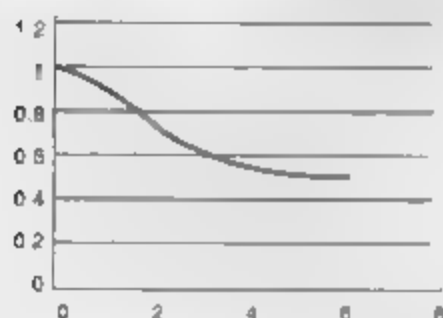


图5 区域评价因子 $F$ 走势图



图6 全城范围内新增平台的平台设置图

表7: 各城区应增加平台的节点统计表

城区	增加平台的节点	增加平台个数
A	28 36 48 57 81	6
B	123 150 159	3
C	195 317 354 242 240 246 261 301 288 280 27 293 206 211	11
D	331 339 357 362 369	5
E	388 391 407 474 413 452 418	7
F	487 495 508 510 518 524 528 512 561 574 578	11

### 3.4.4 综合分析各城区交界区域的交巡警平台设置

分×考虑时,认为每个城区×警力只负责自己城区内的节点,而在实际情况中,警力是受到整个城市指挥系统的调配的。对交巡警平台警力进行跨城区抽调时,采取“就近原则”。因此我们对各个城区交界处的警力分布进行进一步调配。主要考虑城区交界处警力的合并。

**第1步** 对全市交界处的平台和节点进行搜索,找出交界区域。

**第2步** 将交界处看做是一个区,对其采用前节模型的做法,然后计算出增加的平台个数及其位置。

### 第3步 对交界区域的交巡警平台设置进行优化

利用3.1和3.3的评价模型,对交界区域的交巡警平台分布进行评价,如图7所示。得H,结果不必需要新增平台,因此采取撤销低 $\times$ 数并交巡警平台的策略,对交界区域的平台进行优化,撤销的交巡警平台编号为48,100,372。

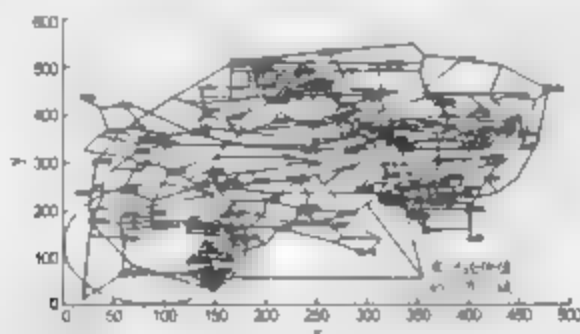


图7 有讨论价值的交界区示意图

### 3.5 最佳围堵方案

在讨论围堵方案,我们先来建立如下两个包围圈

**城市包围圈** 即封锁圈。在接到报案,对于城市进行封锁

**内包围圈** 是追捕圈。计算罪犯从案发地点开始发散可能逃离的范围,是一个圆的包围圈去将罪犯围住,然后将罪犯控制在包围圈内

**外包围圈的封锁** 根据追捕模型建立外包围圈,经过计算,对全市进行封锁的时间为12364min,如图8所示

**内包围圈的封锁** 为减少市民的危害,提高工作效率。应尽量将罪犯限制在小范围内,尽快地将其抓住。



图8 市区外圈的封锁方案

小范围封锁的具体算法如下:

初始化 $t = 0$ ,  $Time_i$ 为从报案开始经过的时间

步骤1:  $t = t + 0.1$ , 计算 需要封锁的点的集合  $S$

步骤2: 对于  $S$  中点, 检查在  $H$  可范围内分步  $q$  以有哪些巡警站 能覆盖到,

步骤3: 由匈牙利算法判断是否加入  $n$  个  $q$  能, 是否成功, 实现了封锁 否则返 步骤1

对于所给逃犯,  $t' = 3 \text{ min}$  再从犯罪人的心理考虑 其会在尽量避免被父巡警发现的情况下 尽量提高车速 经查阅资料 得知各城市车速基本为  $60 \text{ km/h}$ , 所以假设逃跑犯罪嫌疑人 的逃跑速变为  $60 \text{ km/h}$ , 假设警车的速度也是  $60 \text{ km/h}$ , 然后利用计算机进行模拟, 可得: 经过  $6.2 \text{ min}$  时 可将该犯封锁住, 具体封锁方案如图9所示, 其中图1红色箭头  $a \rightarrow b$  表示父巡警从  $a$  点发到  $b$  点封锁路线, 红色阴影表示在封锁区域 蓝色的边表示割边, 红色的边表示 可以完全覆盖的点

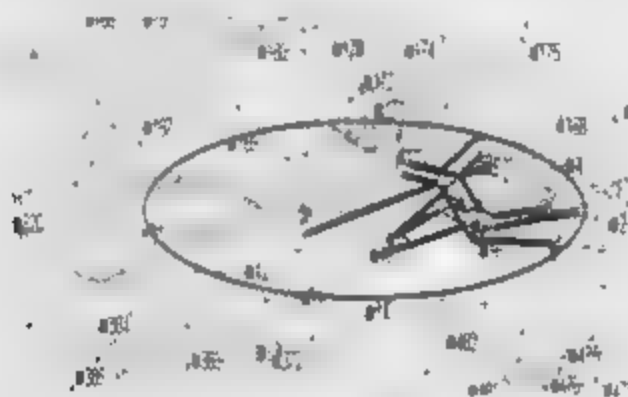


图9 小范围包围图

## 参考文献:

- [1] 余松溪, 余川友. 运筹学简明教程[M]. 北京: 高等教育出版社, 2006.  
Qin Y Y, Yu C Y. A Short Course on Operations Research[M]. Beijing: Higher Education Press, 2006.
- [2] 张莹. 运筹学基础[M]. 北京: 清华大学出版社, 2010.  
Zhang Y. The Basis of Operations Research[M]. Beijing: Tsinghua University Press, 2010.
- [3] 殷剑宏. 图论及其应用[M]. 合肥: 中国科学技术大学出版社, 2003.  
Yin J H, Wu K Y. Graph Theories and its Applications[M]. Hefei: University of Science and Technology of China Press, 2003.
- [4] 卜月华. 图论及其应用[M]. 南京: 东南大学出版社, 2000.  
Bu Y H. Graph Theories and its Applications[M]. Nanjing: Southeast University Press, 2000.
- [5] 胡伯涛. 最小割模型在信息学竞赛中的应用[M]. 北京: NOI 冬令营论文集, 2007.  
Hu B T. Applications of Minimum Cut Model in Informatics[M]. Beijing: NOI Winter Camp, 2007.



## Setting and Scheduling of Traffic and Patrol Police Service Platform

XING Zhong-yang, ZENG Fan-zheng, ZHANG Jiang-hui

College of Opto-Electronic Science and Engineering, National University  
of Defense Technology, Changsha 410073)

**Abstract:** In this paper, we investigate the scheduling and setting of the traffic and patrol police service platforms, aiming to optimize the patrol time and balance the workload for each platform. The Floyd algorithm is used to attain the shortest route, and then the workload of each platform is optimally compromised. Accordingly, each of these optimal administration area can be adaptively allocated and the location of added platforms can be generated, that is, 28, 36, 48, 57, 81. Besides, for quick blocking of the 13 traffic arteries, the Hungarian algorithm is used to calculate the shortest time for a complete blocking, namely, 8.015 minutes. As for the adjustment of the platforms, we propose the evaluation function  $F'$  including the expected workload, the variance, and the longest patrol time. We finally evaluate the optimal scheme for besieging suspects within the blockade circle.

**Keywords:** shortest path, cover within 3 minute, traffic and patrol police service platform



交巡警服务平台, 然后以3分钟出警时间为限, 以出警次数最多的最小或出警次数方差最小建立模型2.1和2.2如下。

### 模型2.1

$$\min k \quad (1)$$

$$\text{s.t.} \quad \sum_{j=1}^{92} x_{ij} p_j < k, \quad i = 1, 2, \dots, 20, \quad (2)$$

$$x_{ij} t_{ij} \leq 3, \quad i = 1, 2, \dots, 20, \quad j \in \{1, 2, \dots, 92\} \setminus \{28, 29, 38, 39, 61, 92\} \quad (3)$$

$$\sum_{i=1}^{20} x_{ij} = 1, \quad j = 1, 2, \dots, 92, \quad (4)$$

$$x_{15,28} = x_{15,29} = x_{16,38} = x_{2,39} = x_{7,41} = x_{20,92} = 1 \quad (5)$$

$$x_{ij} = 1 \quad \forall i = j, \quad j = 1, 2, \dots, 20 \quad (6)$$

$$x_{ij} \text{ 为 } 0 \text{ 或 } 1, \quad i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 92, \quad (7)$$

其中 $i$ 表示A区交巡警服务平台的位置标号, $j$ 表示A区交通网络上路口节点的标号, $x_{ij}$ 为0-1决策变量,若第 $i$ 个交巡警服务平台服务第 $j$ 个路口节点则为1,否则为0; $k$ 表示平台最大出警次数, $p_j$ 表示第 $j$ 个路口节点的案发次数。(1)式为模型的目标函数,其含义是极小化平台上最大的出警次数,该目标主要是考虑出警次数的均衡性。(2)式表示每个平台的出警次数不超过 $k$ ;(3)式表示交巡警要在3分钟内从其所在平台到达其管辖的路口节点(除预先分配的路口节点以外);(4)式表示每个路口节点都有平台管辖;(5)式表示预先分配的路口节点被距离各自最近的服务平台所管辖;(6)式表示位于交巡警服务平台的路口节点都由该服务平台自己来管辖。

利用Lingo软件求解模型2.1可得所需结果,具体路口节点分配方案略。

### 模型2.2

$$\min \sum_{i=1}^{20} \left( \sum_{j=1}^{92} x_{ij} p_j - \bar{p} \right)^2$$

$$\bar{p} = \sum_{j=1}^{92} p_j / 20,$$

其他约束条件同模型2.1中的(3)-(7)式。

其中 $\bar{p}$ 表示出警次数的均值。

同样用Lingo软件求解模型2.2可得所需结果,具体路口节点分配方案略。

比较模型2.1和模型2.2的结果可知(具体细节略去),模型2.2求得方案出警次数均衡性更好。

### 2.3 第二问的解答

根据最短时间矩阵,提取出20个交巡警服务平台到出入A区的13个路口节点的最短时间矩阵 $(t_{ij})_{20 \times 13}$ ,下面建立模型2.3。

### 模型 2.3

$$\min T \quad (8)$$

$$\text{s.t.} \quad \sum_{i=1}^{20} x_{ij} = 1, \quad j = 1, 2, \dots, 13, \quad (9)$$

$$x_{ij} t_{ij} \leq T, \quad i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 13, \quad (10)$$

$$\sum_{j=1}^{13} x_{ij} \leq 1, \quad i = 1, 2, \dots, 20, \quad (11)$$

$$x_{ij} \text{ 为 0 或 1, } i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 13, \quad (12)$$

其中  $T$  为实现全封锁所需最长时间

Lingo 求解可得实现全封锁所需最长时间为 8 015 457 分钟。具体封锁方案略。

注 1 对模型 2.3 我们可进一步改进使得实现全封锁所用的总出警时间最小。模型 2.4 得到的  $T$  值作为约束。以实现全封锁所需的总出警时间最小为目标再建一个 0-1 规划模型进行求解即可。

### 2.4 第三问的解答

在给  $A \times$  增加服务平台时, 应满足出警时间最短和出警次数均衡这两个原则。出警时间最短要求巡逻警在 3 分钟或更短的时间内赶到案发现场, 出警次数均衡原则要求每个平台出警次数尽量均衡。据此建立模型 2.4 及模型 2.5 如下。

#### 模型 2.4

$$\min k \quad (13)$$

$$\text{s.t.} \quad \sum_{j=1}^{92} x_{ij} p_j \leq k, \quad i = 1, 2, \dots, 20, \quad (14)$$

$$2 \leq \sum_{i=21}^{92} m_i \leq Z, \quad Z = 2, 3, 4, 5, \quad (15)$$

$$\sum_{i=1}^{92} x_{ij} = 1, \quad j = 1, 2, \dots, 92, \quad (16)$$

$$x_{ij} t_{ij} < 3, \quad i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 92, \quad (17)$$

$$x_{ij} < m_i, \quad i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 92, \quad (18)$$

$$m_i = 1, \quad i = 1, 2, \dots, 20, \quad (19)$$

$$x_{ij} = 1 \quad \text{若 } i = j, \quad i, j = 1, 2, \dots, 20, \quad (20)$$

$$x_{ij} \text{ 为 0 或 1, } i = 1, 2, \dots, 20, \quad j = 1, 2, \dots, 92, \quad (21)$$

$$m_i \text{ 为 0 或 1, } i = 21, 22, \dots, 92, \quad (22)$$

其中  $m_i$  表示是否设置第  $i$  个路口节点为交巡警服务平台, 若是则为 1, 否则为 0,  $Z$  表示增加平台数量的上限值。标为极小化平台总最大出警次数。(15) 式表示从标号为 21 至 92 的路口节点中选出 2 至 5 个设为服务平台。

### 模型 2.5

$$\min \sum_{i=1}^{92} m_i \left( \sum_{j=1}^{92} x_{ij} p_j - \bar{p} \right)^2 \quad (23)$$

$$\bar{p} = \sum_{j=1}^{92} p_j / \sum_{i=1}^{92} m_i \quad (24)$$

$$\text{其他约束条件同模型 2.4 (即 (15)~(22) 式)} \quad (25)$$

其中  $p_j$  表示  $j$  路警次数均值, 标为极小化出警次数方差。

用 Lingo 软件求解模型 2.4。将  $Z$  设定为 2.3 时, 无解。将  $Z$  设定为 4.5 时, 有解。将  $Z$  设定为 5, 新增平台设置在标号为 21、29、40、48 和 88 的路口节点。将  $Z$  设定为 4, 新增平台设置在标号为 29、39、48 和 88 的路口节点。通过比较(具体细节略去), 我们发现多增设一个平台有利于警次数均衡性, 有利于距离改善, 多增设的那个平台只是分担了那些非常繁忙的平台的工作。效果良好。反而增量、财政负担。在这种情况下, 一般新增 4 个平台比较合适。当然在财政充裕的情况下增设 5 个平台也是可以的。

注 2。与模型 2.4, 我们进一步改进使得所求总出警时间最小。模型 2.4 得到的最优值作为约束, 以总出警时间最小为目标建立 0-1 规划模型进行求解。与类似的方法和, 模型 2.5。

## 3 问题二模型的建立与求解

### 3.1 第一问的解答

#### 1) 评价指标体系的建立

针对全市的具体情况, 为准确地评价交巡警服务平台设置方案的合理性, 首先得建立合理的评价体系。从交巡警服务平台的原则和任务出发, 本文选取如下 3 个指标。

#### 指标 1: 出警时间

出警时间是指从交巡警服务平台到达其所管辖的路口节点所需的时间, 该指标越小越好。

#### 指标 2: 出警次数均衡性

该指标是使各交巡警服务平台的出警次数尽量均衡, 可以用极小化平台出警次数的方差或平台总出警次数最多的最小表示, 即  $\min\{\max_i \sum_j x_{ij} p_j\}$ 。

#### 指标 3: 总出警时间

总出警时间是指所有交巡警服务平台赶赴案发点所花时间之和, 即  $\sum_i \sum_j x_{ij} t_{ij} p_j$ 。

在实际中, 出警时间最重要, 出警次数均衡性次之, 总出警时间作为最弱。

#### 2) 合理性评价

为评价现有方案, 我们先对交巡警服务平台采用重设平台的方式进行优化, 通过比较现有方案和优化后方案的指标, 确定现有方案是否合理, 是否需要改进。为此建立优化模型如下。

### 模型 3.1 根据指标1建立最长出警时间最短模型

模型如下所示

$$\min T \quad (26)$$

$$\text{s.t.} \quad \sum_{i=1}^{a_z} x_{ij} = 1, \quad j = 1, 2, \dots, a_z, \quad (27)$$

$$x_{ij} t_{ij} \leq T, \quad i = 1, 2, \dots, a_z, \quad j = 1, 2, \dots, a_z, \quad (28)$$

$$x_{ij} \leq m_i, \quad i = 1, 2, \dots, a_z, \quad j = 1, 2, \dots, a_z, \quad (29)$$

$$\sum_{i=1}^{a_z} m_i = b_z, \quad (30)$$

$$x_{ij} \text{ 为 } 0 \text{ 或 } 1, \quad i = 1, 2, \dots, a_z, \quad j = 1, 2, \dots, a_z, \quad (31)$$

$$m_i \text{ 为 } 0 \text{ 或 } 1, \quad i = 1, 2, \dots, a_z, \quad (32)$$

其中  $z = 1, 2, \dots, 6$  分别表示该市 A、B、C、D、E、F 六个区,  $b_z$  表示第  $z$  区现有的平台数量,  $a_z$  表示第  $z$  区的路口节点数. 约束条件 (30) 表示新设置的平台总数和现有方案相等.

### 模型 3.2 根据指标2建立出警次数均衡性模型

以平台中出警次数最多的最小为目标, 并把模型 3.1 中得到的  $T$  值作为约束, 建立模型如下

$$\min k$$

$$\text{s.t.} \quad \sum_{j=1}^{a_z} x_{ij} p_j \leq k, \quad i = 1, 2, \dots, a_z,$$

其他约束条件同模型 3.1 中的 (27)~(32) 式.

或以平台中出警次数的方差为目标, 用模型 3.1 得到的  $T$  值作为约束, 建立模型如下

$$\min \sum_{i=1}^{a_z} m_i \left( \sum_{j=1}^{a_z} x_{ij} p_j - \bar{p}_z \right)^2$$

$$\text{s.t.} \quad \bar{p}_z = \sum_{j=1}^{a_z} p_j / \sum_{i=1}^{a_z} m_i,$$

其他约束条件同模型 3.1 中的 (27)~(32) 式.

### 模型 3.3 根据指标3建立总出警时间最少模型

以总出警时间最少为目标, 并把模型 3.1 和模型 3.2 得到的  $T$  值和  $k$  值作为约束, 建立模型

如下

$$\min \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} x_{ij} t_{ij} p_j$$

$$\text{s.t.} \quad \sum_{j=1}^{a_2} x_{ij} p_j < k, \quad i = 1, 2, \dots, a_1,$$

其他约束条件同模型3.1中的(27)~(32)式.

用 Lingo 软件求解 得到需求结果, 见表1及表2. 比较表1和表2, 发现+重设平台方式得到的优化方案在出警时间上有了很大的改善, 而其他的指标变化不是很明显. 这说明现有方案+增合理需要改进. 在求解中, 出警时间起着约束性的作用, 这是设计改造警服务平台的第一原则. 只有在满足该原则上, 考虑+警次数总衡性才有意义.

表1: 现有方案的各个指标

区号	A	B	C	D	E	F
出警时间	3.7005	4.470312	6.8605	16.06282	19.105	8.47985
最多出警次数	7	8.4	13.1	7.7	8.4	10.8
出警次数方差	89.9	0.0175	10.87035	0.1133	0.5557	26.65
总出警时间	142.90	97.01	403.307	176.3078	379.14	293.16

表2: 以重设平台方式得到的优化后方案的各个指标

区号	A	B	C	D	E	F
出警时间	3.4713	3	4.6	6.8529	5.0221	7.6932
最多出警次数	7	8.4	13.1	8.5	8.5	10.8
出警次数方差	1.9949	0.0275	8.4634	4.8911	1.7637	3.5147
总出警时间	145.6697	73.97	344.54	121.55	306.21	223.32

### 3) 优化后调整方案

观察表2可知, A区和B区出警时间基本上在3分钟之内, 增设平台又需要成本, 所以本文认为A区和B区通过重设平台就能得到较合理的方案. 对于C、D、E、F区则需要用增设平台方式来优化, 这只要在模型3.1~模型3.2和模型3.3中逐步增大 $b_2$ 的值即可. 得到的评价结果见表3. 比较表2和表3, 发现用增设平台的方式得到的优化方案在出警时间上有了较大的改善, 而其他的指标也有一定程度的改善. 具体调整方案见表4.

在财政紧张的情况下, 用重设平台方式也能得到较好的结果, 其调整方案见表5.

### 3.2 第二问的解答: 拥堵方案(略)

表3 以增设平台方式得到的优化后方案的各个指标

区号	C	D	E	F
出警时间	4	4 6057	4 2202	4 1012
最多出警次数	8.4	7.5	8	8
出警次数方差	6.4316	6.7758	2.9893	6.0333
总出警时间	274.89	130.44	247.53	176.26

表4 以增设平台方式得到的优化后方案

区号	各交叉警务服务平台位置所在节点标号
C	166, 167, 171, 175, 179, 186, 191, 201, 210, 219, 224, 225, 227, 244, 247, 254, 261, 269, 271, 276, 287, 291, 296, 311, 316
D	321, 324, 325, 330, 334, 337, 343, 345, 347, 361, 362, 370
E	373, 375, 379, 385, 387, 388, 391, 398, 403, 408, 419, 441, 446, 453, 457, 458, 466, 468
F	475, 478, 479, 482, 483, 485, 492, 507, 521, 525, 540, 548, 553, 572, 567, 575, 577, 582

表5 以重建平台方式得到的优化后方案

区号	各交叉警务服务平台位置所在节点标号
A	3, 4, 5, 9, 10, 17, 18, 20, 21, 24, 27, 29, 31, 36, 40, 48, 59, 69, 76, 89
B	99, 102, 106, 116, 123, 133, 147, 164
C	167, 169, 170, 173, 175, 179, 186, 201, 210, 213, 228, 240, 262, 272, 286, 307, 314
D	320, 324, 326, 329, 334, 341, 361, 367, 170
E	384, 387, 388, 391, 396, 403, 405, 418, 423, 429, 448, 450, 456, 457, 460
F	489, 496, 511, 518, 539, 542, 548, 553, 557, 569, 576

## 参考文献:

- [1] 谢金星, 邢文训. 网络优化(第2版)[M]. 北京: 清华大学出版社, 2009.  
Xie J X, Xing W X, Wang Z B. Network Optimization (2nd Edition)[M]. Beijing: Tsinghua University Press, 2009.
- [2] 姜启源, 谢金星, 叶俊. 数学模型(第3版)[M]. 北京: 高等教育出版社, 2003.  
Jiang Q Y, Xie J X, Ye J. Mathematical Model (3rd Edition)[M]. Beijing: Higher Education Press, 2003.
- [3] 袁新丰, 邵大志, 郁时康. Lingo和Excel在数学建模中的应用[M]. 北京: 科学出版社, 2007.  
Yuan X S, Shao D H, Yu S L. Application of Lingo and Excel in Mathematical Modeling[M]. Beijing: Science Press, 2007.
- [4] 薛定宇, 陈阳泉. 高等应用数学问题的Matlab求解(第2版)[M]. 北京: 清华大学出版社, 2008.  
Xue D Y, Chen Y Q. Matlab Algorithm of Higher Applied Mathematics Problems (2nd Edition)[M]. Beijing: Higher Education Press, 2008.



## Modeling the Setting and Scheduling of Traffic and Patrol Police Service Platforms

WANG Wei, LI Xin, ZHANG Jian-peng

Advisor: LUO Wen-chang

(Faculty of Science, Ningbo University, Ningbo 315211)

**Abstract** In this paper, we consider the problem of setting and scheduling of traffic and patrol police service platforms. First, we obtain the shortest path matrix of nodes by programming with the Matlab software, and then transform it into the shortest time matrix of nodes. For the first question of Problem one, we pre-assign 6 nodes that the police cannot arrive at within 3 minutes and then propose the corresponding 0-1 programming models with the goals to minimize the maximum workload of police and the variance of workload of police under the constraint that the responding time of police is no more than 3 minutes, respectively. By comparing these results, we show that the latter is better in compromising the workload of police. For the second question of Problem one, we present a 0-1 programming model with the goal to minimize the longest responding time of police subject to blockading at the 13 access arteries. By solving the proposed model, we show that the longest blockading time is 81' 55 minutes. For the third question of Problem one, we derive the corresponding 0-1 programming models with the goals to minimize the maximum workload of police and the variance of workload of police under the constraint that the responding time of police is no more than 3 minutes. After the results are compared, for the first question of Problem two, we propose three indexes, i.e. the responding time of police, the workload of police and the total responding time of police, to evaluate the reasonableness of the present schemes, respectively. According to the three indexes, we propose the corresponding 0-1 programming models step by step to evaluate the reasonableness of present schemes. Based on the three indexes, we give the evaluation results for the six districts (A, B, C, D, E, F). By resetting the platforms or adding the new platforms, we give the adjustment schemes for the six districts (A, B, C, D, E, F). For the second question of Problem two, we first get the nodes of each level from P node to arrive at by programming with Matlab. Then, by using the idea of gradually giving back to constrain the suspects, we propose the containment strategy. The suspects can be contained within 7 362 minutes by using the new strategy.

**Keywords** setting and scheduling of platforms; 0-1 programming; responding time of police; workload of police



此问要求路口节点能够快速从所属的平台到达,在不考虑均衡各平台工作量的情况下,将在下文讨论),需将各路口分配给离它最近的平台管辖,利用Floyd算法为 $A \times 20$ 个交巡警平台合理分配了管辖范围

### 4.2 封锁方案的制定

此问要求调度20个平台的警力来封锁A区的13个出口,每个平台的警力只能封锁一个出口,而每个出口也需要一个平台,由于这种一一对应的关系,很自然地将问题化归为典型的匹配模型,将每个封锁方案看作一个向量 $(x_1, x_2, \dots, x_{13})$ ,  $x_i \neq x_j$ , 平台 $x_i$ 用来封锁出口 $y_i$ , 优化指标为:  $\min \max m(x_i, y_j), i = 1, 2, \dots, 13$

像这样定义的匹配模型难以直接求得最优解,因而需要对问题进行转化,考虑先将问题转化为判定性问题,即对给定的时间阈值,能否找到一种匹配方案成功封锁,然后考虑缩小时间,逼近最优解,设定了时间阈值后,网络流模型能很好地解决判断是否能在该时间内封锁所有路口这一问题,网络流模型的重要平衡条件能够保证每个平台只封锁一个路口,而最大流就是最多封锁的路口数,将此算法记为算法一,具体的执行步骤如下

算法一:

步骤1 从源到所有平台 $x_i$ 连一条流量上限为1的边

步骤2 从所有要封锁的出口 $y_j$ 到汇连一条流量上限为1的边

步骤3 对任意一对平台 $x_i$ 和出口 $y_j$ ,若它们之间的最短时间 $t_{ij}$ 小于等于给定的阈值,则从平台 $x_i$ 到 $y_j$ 连一条流量上限为1的边;

步骤4 计算模型的最大流

步骤5 若流量等于出口数,则所有出口在规定时间内均被封锁

用二分法求得满足条件的最小时间阈值为8.0157 min, 相对应的警力调度方案如表1

表1: 快速封锁警力调度方案

路口序号	12	14	6	21	22	23	24
平台编号	A10	A16	A2	A11	A12	A14	A13
路口序号	28	29	30	38	48	62	
平台编号	A15	A7	A5	A1	A4	A3	

### 4.3 服务平台的增设

此问要求根据现有交巡警服务平台的工作量不均衡和有些地方报警时间过长的实际情况增设服务平台,建立以最长报警时间及工作量均衡度为目标的双目标规划模型并采用兼容分层求解法求解,最终确定为A区增加4个平台,位置分别为28、40、48、91号节点

## 5 问题二的模型

### 5.1 全市平台设置的合理性

此问要求分城考虑平台设置方案的合理性,建立综合评价模型,提出出警时间达标率、平均超出时间以及平均工作量这三个评价指标,进而用夹角余弦法确定指标的权重,经求解比较后,认为A、B两区的平台设置方案相对合理,其余四区的方案相对不合理,其中D、E两区主要是位置安排不合理,C、F两区主要是警力不足,针对不合理各区的特定原因,我们给出了相应的解决方案<sup>[2,3]</sup>.

## 5.2 抓捕犯罪嫌疑人

关于嫌犯, 我们首先假设其踪迹在案发后不可追踪, 即3分钟内嫌犯可以跑到路程范围允许的一切地方, 而警方对此没有具体信息。其次, 假设嫌犯十分狡猾, 如果有任意一条逃离市区道路没有在合适的时间内被封锁, 抓捕工作都会失败。最后, 假设嫌犯的车速可快可慢, 可掉头、转弯, 但最大车速固定, 记作 $v$ 。

抓捕工作分两步完成。首先迅速建立包围圈将嫌犯困住, 使嫌犯无法逃出该市。若无法建立这样一个包围圈, 则抓捕工作失败。其次在包围圈内搜捕嫌犯, 搜捕工作的基本前提是不能放跑嫌犯。目标为在尽量短的时间内抓获嫌犯。若效率相当, 则使用警力的越少越好。

### 5.2.1 围堵模型

在寻找判据是否能围堵成功之前, 首先介绍两个定义。若 $u, v$ 属于点集 $S$ ,  $u, v$ 间有邻点均属于 $S$ , 则称 $u, v$ 为集合 $S$ 的内点。反之, 则称其为 $S$ 的边界点。将点集 $u$ 记作 $V_1$ , 边界集记作 $V_2$ ,  $S = V_1 \cup V_2$ ,  $V_1 \cap V_2 = \emptyset$ 。若 $u, v \in V_2 \cap S$ , 将 $u, v$ 间有邻点加入 $S$ 称为 $S$ 的次增<sup>[1]</sup>, 此时 $v \in V_1$ 。如果我们能够找到一个集合 $S$ , 包含嫌疑人的位置 $u$ , 且集合 $S$ 的所有边界点警察都能赶在嫌犯之前到达, 则围堵成功。否则, 围堵失败。我们希望在最小的地理范围内进行围堵。图1的围堵模型如图1所示。具体步骤如下。

**步骤1** 将嫌犯在案发3分钟内可能到达过的位置 $u$ 存路, 并置集合 $u$ 为初始集, 记作 $S_0$ 。

**步骤2** 将 $S_0$ 的所有边界点 $v$ 次增<sup>[1]</sup>, 得到新的点集记作 $S_1$ 。显然 $S_0 \subset S_1$ ,  $S_0$ 中所有的点为 $S_1$ 的内点,  $S_1$ 满足连续围堵条件的最小集合。嫌犯若要逃离, 必须经过 $S_1$ 的某个边界点 $v_2 \in S_1$ 的次增<sup>[1]</sup>, 令 $S_1$ 增<sup>[1]</sup>成 $S_2$ 。警方只要赶在嫌犯之前占领这些边界点, 则围堵成功。

**步骤3** 若不能在上面为 $S_2$ 围堵成功, 则选择其中一个边界点进行增<sup>[1]</sup>, 再进行判断。在边界点选择上, 我们基于贪婪法, 选择嫌犯到达时间和最近 $v_2$ 到达时间差最大的边界点。基本想法是增广最难包围的边界点。

对于 $v_2$ 的次增<sup>[1]</sup>可以采用算法2进行判断是否可以包围, 当 $v = 1$ 时的包围圈如图2所示, 具体步骤如下:

**步骤1** 从源到所有中间点连一条流量上限为1的边。

**步骤2** 从所有边界点 $v_2$ 连一条流量上限为1的边。

**步骤3** 若 $v_2$ 到边界点 $v_1$ 的时间 $mu(v_2, v_1) < mu(v_2, v_3) + 3$ , 则从中间点 $v_2$ 向边界点 $v_1$ 连一条流量上限为1的边。

**步骤4** 计算在给定的包围圈下模型的最大流。

**步骤5** 若流量等于边界点数目, 则围堵成功。

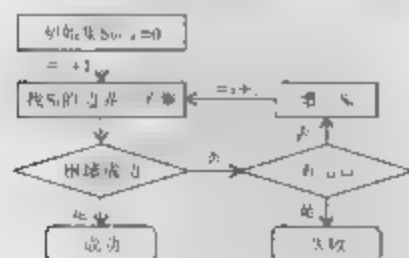


图1 围堵模型

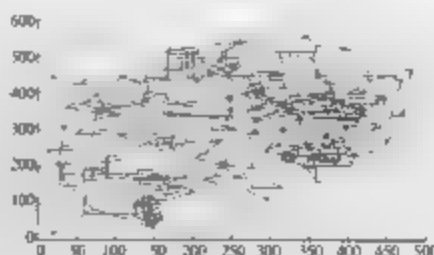


图2 当 $v=1$ 时的包围圈示意图

我们将以将围堵结果进行灵敏度分析,在参数 $\lambda$ 取不同值的情况下,我们得到不同的包围圈及其所包含节点数,见表2。当嫌疑人的最大车速与警车速度相比较小时,用来围堵嫌疑人的包围圈所含节点数对速度的变化比较敏感,当嫌疑人车速与警车速度相当时,包围圈所含节点数表现出相当良好的稳定性,而当嫌疑人车速比警车速度快到一定程度时,我们发现在此问题中已经无法围堵嫌疑人。

表2 包围圈所含节点数与嫌疑人最大车速的关系

$v$ (km/min)	0.5	0.6	0.7	0.8	0.9	1
节点数	9	12	20	22	20	21
$r$ (km/min)	1	1.2	1.3	1.4	1.5	$> 1.5$
节点数	22	21	21	22	22	不可

### 5.2.2 收缩模型

将每个平面的警力称为一支交警队。围堵成功时,逐步缩小包围圈进行抓捕,初步考虑,所有参与围堵的交警队和包围圈内部平面的交警队联合进行。当警车与嫌疑车辆在某道路上的节点相遇时,抓捕成功。若能每次成功减少包围圈的一个边界点,不放跑嫌疑人,则在有限步操作后必能抓捕成功,具体流程如图3。定义一个边界点邻接的内点数量为该边界点的内点度。定义一个边界点邻接的内点与边界点的数量总和为 $1$ 的邻点度。首先考察最外层的包围圈(即成功围堵的包围圈),所有内点度为 $1$ 的边界点必能被收缩至与之邻接的唯一内点,即将该边界点的警力调到邻接的唯一内点上。然而,该操作过后,再直接收缩内点度为 $1$ 的边界点时,可能会放跑嫌疑人,原因是最外层包围圈的边界点之间不可能存在嫌疑人,但是收缩之后的包围圈的边界点之间有可能存在嫌疑人,如图4中右图。



图3 缩小包围圈的步骤

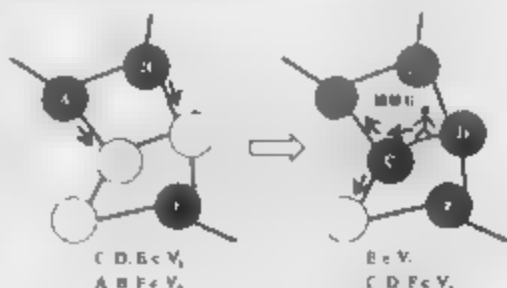


图4 边界点收缩示意图

由此我们得出收缩操作的第一个约束条件——邻点度约束。在前一次收缩完成后,若要再收缩某个边界点,其上的交警队数目必须大于等于其邻点度。由于我们在追捕环节的目标是化最少时间抓到嫌疑人,因此需要一个指标来引导我们的收缩操作。现在引入准备时间的概念。准备时间指的是点 $i$ 最早可以开始收缩的时间,记作 $pre(i)$ 。首先设包围圈中所有 $i$ 的准备时间为0。选定收缩的点 $i$ 以后,将其警力分配到所有邻点上,不得遗漏,因为在处理最外层包围圈后,嫌疑人可能位于两个边界点之间。邻点 $j$ 的准备时间更新为

$$pre(j) = \max \{ pre(i) + t(i, j), pre(j) \}$$

通过这个操作,虽然我们每次只收缩一个边界点,但实际上它们是可以同时进行的。每次只收缩一个点相当于给定一个收缩顺序,相当于每次交警队到达新的地点时就原地待命,直到下一步指示可以继续缩小包围圈。在收缩过程中互不影响的同时进行。如图2所示的局部包围圈,边上的数字代表路警。A点的准备时间为5,假设它可以收缩,则B和C的准备时间更新为7和8,初始为0。紧接着收缩B点,D的准备时间更新为8。最后收缩C点,将D的准备时间更新为11。虽然每次只收缩了一个点,但是实际模拟的是A的两批警力同时出发,并在D集合待命。

对于任意一种收缩顺序,如果最后其能收缩为一个点,则最后一个点的准备时间即为搜捕时间,优化目标为 $\min pre(k)$ ,对于任意收缩顺序,其中 $k$ 是最后的收缩点。

于此搜捕模型没有考虑没有可收缩边界点的情形。在编程求解的过程中,这个问题也确实发生了,当迭代到某一步, $k_2$ 中没有任何一个点可以收缩,则搜捕失败。考虑到 $G$ 是一张平面图,所以每个点的邻接点数都不太大,大多数小于5;相反,随着包围圈的缩小,警察越来越集中,收缩点应该在上述失败。这个矛盾的产生是因为在收缩的过程中,收缩点有可能将包围集撕裂。比如在图5中,若A和B收缩完毕, $D$ 也成为了边界点,它可以收缩。当已收缩时一部分警力分配给了C,并将C孤立。C只能收缩至自身,其上含有的所有警力将无法继续参加包围,形成浪费。基于以上分析,我们得出收缩操作的第一个约束条件:连通性约束。每次收缩的边界点都不能破坏包围集的连通性。连通性是指集合内的任意两点,可以只通过集合内部的点与达。对于我们的优化目标 $\min pre(k)$ ,采用贪婪法选取收缩顺序,每次选择准备时间最小的可收缩边界点,将其警力分散到所有邻点。若警力不能被邻点数整除,则将余数任意分配至某邻点。

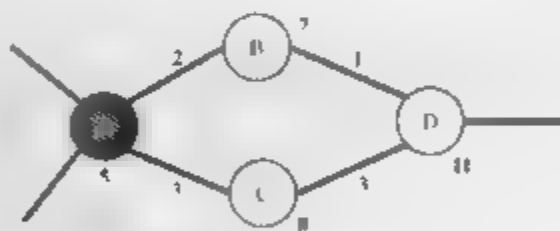


图5: 准备时间示意图

### 5.2.3 结果与优缺点分析

- 1) 在嫌犯车速为1时,抓捕过程共耗时56分钟。
- 2) 若把包围集内交警队数目等于点数的情况视作搜捕成功,则整个过程只需29分钟,当包围集所含点数为26时迭代结束。
- 3) 模型的显著优点是推广性强,因为它并不依赖于点P,无论嫌犯的初始位置在哪里,模型都能尽快围堵或判定围堵失败。
- 4) 模型缺点:是没有考虑最外层包围集以外的其他警力,可能会导致搜捕时间偏长,是在增和收缩的过程中都采用了贪婪法,解的质量可能会不高。
- 5) 改进方向:在收缩或增过程中,加入一些评估函数。但因为时间有限我们未能找到比贪婪法更好的智能搜索,实为遗憾。

## 参考文献:

- [1] 张军. 计算智能[M]. 北京: 清华大学出版社, 2009.  
Zhang J. Computational Intelligence[M]. Beijing: Tsinghua University Press, 2009.
- [2] 吴礼斌. 数学实验与建模[M]. 北京: 国防工业出版社, 2007.  
Wu L. B. Mathematics Experiment and Modeling[M]. Beijing: National Defence Industry Press, 2007.
- [3] 姜启源. 数学模型(第3版)[M]. 北京: 高等教育出版社, 2009.  
Jiang Q. Y. Mathematic Model (3rd Edition)[M]. Beijing: Higher Education Press, 2009.

## The Optimal Setting and Scheduling of the Traffic and Patrol Police

ZHANG Xiao, TIAN Zhi-xiang, FU Jian-long  
Advisor: Instructor Group

School of Finance, Shanghai University of Finance and Economics, Shanghai 200433)

**Abstract:** This paper first employs the Floyd algorithm to preliminarily allocate the police stations, and builds the network-flow model to establish a scheme to blockade an area. Second, a multi-objective programming model is used to evaluate the current allocation of the police resource and present the corresponding improvement. Finally an encircle-catch model is set up. The network flow model is utilized again to make different encircling plans according to the suspect's speed, and on the basis of the greedy algorithm the encirclement under two constraints is constructed to catch the suspect within one hour.

**Keywords:** network flow model; multi-objective programming model; simulated annealing algorithm; comprehensive evaluation model





3) 建立以下0-1规划模型, 寻求以  $M_2$  为围堵节点的最佳警力配置方案

$$\begin{aligned} \min z &= \max_{j,k} \{d_{jk}a_{jk}\} \\ \text{s.t. } &\begin{cases} \sum_k a_{jk} = 1, \\ \sum_j a_{jk} \leq 1, \\ d_{jk}a_{jk} \leq d_{32j} - v - 3\text{min}, \\ a_{jk} = 0, 1, \quad j \in M_2, \quad k \in M_3. \end{cases} \end{aligned}$$

1.1. 1 称为最大围堵距离最小(等价于最大围堵时间最短)。决策变量  $a_{jk}$  指节点  $j$  是否由平台  $k$  去封堵,  $\sum_k a_{jk} = 1$  指平台  $k$  至多围堵一个节点。其他约束含义清楚, 不再赘述。

若该问题有最优解, 则下一步继续。否则(等价于没有可行解), 表明以  $M_2$  作为围堵节点时, 任何警力配置下都不能保证将罪犯围堵, 所以需要修改集合  $M_1$ , 回到下一步继续。

寻找修改  $M_1$  的最佳方法, 可能需投入大量计算时间。所以近似计算警力配置方案(值), 不去求解 0-1 规划, 并同时对其  $M_1$  进行修改。例如简单地依次对  $j \in M_2$ , 总是在  $M_3$  中寻找离节点  $j$  最近的平台  $k$  去封堵。若在  $j$  处不可以将罪犯嫌疑人围堵掉, 即对任意未分配任务的平台  $k$ ,  $d_{jk} > d_{32j} - v - 3\text{min}$ , 则将  $j$  加入  $M_1$  中, 回到下一步继续。

4) 如果上述优化问题有最优解, 此时以  $a_{jk}$  为围堵方案,  $M_2$  中所有节点都能被围堵掉, 则  $M_2$  为所要围堵节点,  $M_1 + M_2$  为围堵区域。

### 6.2.2 模型的求解

对上述模型在 Matlab 软件中求解, 算法如下:

步骤 1 以  $P$  为出发点, 确定不能封锁的点的集合  $A$  的初值(3 分钟之内的点);

步骤 2 确定与  $A$  相邻的点组成的集合  $B$ ;

步骤 3 考察  $B$  中的所有节点, 如果  $B$  中任一节点不能及时封堵, 则将该节点加入  $A$ , 并跳转到第 1。

步骤 4 如果  $B$  中所有节点都能在罪犯逃离该节点前被及时封堵(本文在编程中采用警力配置就近原则进行封堵), 则  $B$  即为所求封锁节点的集合,  $A+B$  为所求封锁区域。

分别取  $v=60, 70, 80, 90 \text{ km/h}$ , 可求得结果, 如表 1 所示。当  $v=60 \text{ km/h}$  时, 通过设防节点和围捕平台的最佳标记得出围捕时间为 8.9 min。

表 1: 不同逃逸速度下的围捕方案

60	设防节点	3 4 5 6 10 15 16 41 40 63 60 171 173 168 240 230 248 273 370 371 561
	围捕平台	3 4 5 6 10 15 16 17 02 01 168 171 173 175 172 170 167 182 320 321 475
70	设防节点	3 4 5 6 10 15 14 41 40 63 620 173 231 168 243 248 253 273 370 371 488 549 561
	围捕平台	3 4 5 6 10 15 14 17 02 01 168 173 171 175 172 167 169 170 320 321 482 481 475
80	设防节点	3 4 10 15 14 41 40 62 64 233 231 168 230 248 250 254 273 349 370 488 549 558 562
	围捕平台	3 4 10 15 14 17 02 168 01 173 171 175 172 167 182 169 170 320 321 482 481 475 480
90	设防节点	3 4 10 14 41 40 85 64 171 173 190 229 168 248 250 254 273 349 389 488 549 558 562
	围捕平台	3 4 10 14 17 02 20 01 171 173 168 172 175 167 182 169 170 320 321 482 481 475 481

## 7 模型评价(模型 2.2)

优点: 根据犯罪嫌疑人被发现之前所能走到的所有最近点建立集合, 并从这个集合出发搜索所有的“延伸”节点, 并通过延伸出的节点与最近警力点的距离和犯罪嫌疑人距离的差值来找到封锁点。这样所得到的点既可以满足题目给书的要求, 且程序设计相对容易。

缺点: 只能存在警力, 并非最近的调度才是最佳的情况。因为可能会出现更远的警力点到相对远的节点的可能。这样时间会很长。而本身如果将这个警力点分配到相对远的节点, 而更近的警力点到就近的节点, 时间会更短。本文未将其考虑在内。

## 8 附录 2.2 程序<sup>[1,2]</sup>

```
clear; close all; clc
zuo=xlread('C:\Documents and
Settings\Administrator\桌面\cumcm.xls',1,'B2:C583'),
lv=xlread('C:\Documents and
Settings\Administrator\桌面\cumcm.xls',2,'A2:B929'),
dist=zeros(1,1),k=1,
for i=1:928
    dist(1,k)=sqrt((zuo(lv(i,1),1)-zuo(lv(i,2),1))^2
    +(zuo(lv(i,1),2)-zuo(lv(i,2),2))^2),
    k=k+1;
end
dist=dist/10;lu=zeros(582,582);k=1;
for i=1:928
    lu(lv(i,1),lv(i,2))=dist(k);
    lu(lv(i,2),lv(i,1))=dist(k);
    k=k+1;
end
for i=1:582
    for j=1:582
        if lu(i,j)==0
            lu(i,j)=10000,
        end
    end
end
for i=1:582
    lu(i,i)=0,
end
weight=lu,
for k=1:582
    for i=1:582
        for j=1:582
            if weight(i,j)>weight(i,k)+weight(k,j)
```

```

        weight(i,j)=weight(i,k)+weight(k,j);
    end
end
end
tic speed=60;      %小偷速度设定(km/h)
station=xlsread('C \Documents and
Settings\Administrator\桌面\cumcm.xls',3,'B2:B81'),
numstation=length(station), %警力点个数
sp=speed/60; numA=0;
for i=1 582      %确定无法封锁的节点集合A的初值 (P点出发 一分钟之内的点)
    if(weight(32,i)<(3*sp))
        numA=numA+1,
        A(numA)=i;
    end
end
fail=1,times=0,
while((numA<583)&&(fail~=0))
    times=times+1,
    fail=0,numB=0,numD=0,
    B=0,D=0,Dstation=0;
    station(:,2)=1;
    for i=1 928
        if((~isempty(find(A==lv(i,1), 1)))&&(isempty(find(B==lv(i,2), 1)))
            &&(isempty(find(A==lv(i,2), 1))))
            numB=numB+1;
            B(numB)=lv(i,2);
        end
        if((~isempty(find(A==lv(i,2), 1)))&&(isempty(find(B==lv(i,1), 1)))
            &&(isempty(find(A==lv(i,1), 1))))
            numB=numB+1,
            B(numB)=lv(i,1);
        end
    end
end
% B=B(2:numB), %B集合为A集合的相邻节点构成的集合
% numB=numB-1;
for i=1 numB,
    Ddistance=0,
    d1=weight(32,B(i))/sp,%小偷到该节点的最短时间
    for j=1 numstation
        d2(j,1)=weight(station(j,1),B(i)),%各警力点到该节点的最短时间
    end
end

```

```

D=find((d1-d2)*station(:,2)>3),
if isempty(D)
    numA=numA+1; A(numA)=B(i);
    B=[B(1:i-1),B(i+1:numB)];
    fail=1; break,
else
    numD=numD+1,
    for k=1:length(D)
        Ddistance(k)=d2(D(k),1);
    end
    [Dmin,numDmin]=min(Ddistance);
    Dstation(numD)=station(D(numDmin),1),
    station(D(numDmin),2)=0,
end
toc display(A) display(B) C=[A,B]

```

## 参考文献:

- 1 张志涌. Matlab教程[M]. 北京: 北京航空航天大学出版社, 2006.  
Zhang Z Y. Matlab Tutorial[M]. Beijing: Beihang University Press, 2006.
- 2 吴茂. 数学建模及其基础问题的求解[M]. 武汉: 武汉大学出版社, 2006.  
Wang W B. Details, Annotation of Mathematical Modeling and Basic Knowledge[M]. Wuhan: Wuhan University Press, 2006.

# Setting and Scheduling of Traffic and Patrol Police Service Platforms

LIU Jin, YANG Yun, WU Ke

Faculty of Mathematics Statistics, Wuhan University, Wuhan 430072

**Abstract:** Assuming that each police platform can only block one point, we create a dynamic optimization model for the optimal setting and scheduling problem of the traffic and patrol police service platforms under the consideration of the uncertainty of the criminal's fleeing direction. We first create a subset A which consists of the points with less than 3-minute away from the criminal's fleeing start point. Then we formulate a cut-set B which consists of the adjacent points of A, and search and allocate the adjacent points (transforms B the points in cut-set B). If no police platforms can be allocated, we reformulate cut-set B again until an appropriate blocking condition tallies with the requirements. In this paper, we evaluate each nearly optimal blocking strategy when the criminal's speed is 60, 70, 80 or 90 km/h. Specifically, when the criminal's speed is 60 km/h, the fastest blocking time is 8.9 min.

**Keywords:** 0-1 programming; Floyd's algorithm; platform scheduling; best block way

文章编号 1005-3085(2011)07-0192-11

## 天然肠衣搭配优化问题的模型和计算

陆立强<sup>1</sup>, 俞冰清<sup>1</sup>, 范 岸<sup>2</sup>

1. 复旦大学数学科学学院, 上海 200433, 2. 中国银联股份有限公司, 上海 200135

**摘 要** 本文介绍天然肠衣搭配优化问题的建模、问题的分析、整理、求解过程, 讨论了两种解题思路以及相关的数学模型、计算方法以及数值结果.

**关键词** 肠衣搭配; 整数规划; 数学建模

**分类号** AMS 2000: 90B90; 90C29; 90C59

**中图分类号** O221.1

**文献标识码** A

### 1 问题的背景

天然肠衣(以下简称“肠衣”)指的是宰杀的人、小肠经刮制而成的副产品,主要加工香肠、灌肠等食品的外衣.中国加工肠衣历史悠久,是肠衣生产大国,其产量占世界总产量的一分之一,其产量80%以上,年产量达到数百万桶,出口金额为4亿美元左右.欧洲是我国肠衣的主要出口对象,目前50%以上的出口到欧洲,欧洲市场70%左右的肠衣来自于中国.我国肠衣出口主要到美国、加拿大、日本、俄罗斯等国家地区.自1985年对肠衣生产企业实施出口食品卫生注册以来,我国加工肠衣生产企业的整体管理水平得到了很大的提高.我国现有肠衣加工企业200家左右,其中对欧盟注册的就有119家左右,有力地保证了我国肠衣的对外贸易.同时,国内市场对肠衣的需求日益递增趋势,对于肠衣生产加工企业,无论是国际市场,还是国内市场,需求均在不断地增大,商业机会越来越多,竞争也更加激烈.

肠衣加工过程中,首先需要剪裁原材料.由于化学-物理作用,难免会导致材料破损,为而必须将破损的部位剔除.这是造成原材料长度不一的主要原因.首先是捆扎.即将清洗干净的肠衣捆成一捆,规格捆扎成成品.一般要求每捆成品中肠衣的总长度是充足的,所以捆扎过程必须首先测量每根原材料长度,然后将其长短不一的原材料搭配成捆.

1.1 原料肠子匹配.软、滑等特性.由于肠衣加工主要依靠人工,其捆扎环节尤其要求工人眼手快,所谓眼明是指原料长度比较准确,手快则是可以快速判断手中的原料是否可以用.测量过的原料捆扎成捆成品.符合这样要求的另一方面年龄不能太大,否则速度太慢,影响加工速度,导致原料报废,另一方面还得具备一定的经验(否则,无法将长短不一的原料快速捆扎成捆).伴随着行业竞争的加剧,人工成本的增加是一种必然的趋势.

作为“种食品”,一般都是原料当人进厂,当人就要加工完成出厂,不允许将剩余的原材料留存以后使用,因此对于原料的使用率有比较高的要求.与此相对的是,人搭配时,人一般不作整体考虑,只是根据手中现有的部分原料,凭经验和直觉判断是否可以搭配成捆成品,不能考虑到所有情况,无法保证原材料的充分利用.

所以无论从降低工人劳动强度与压缩人力成本,还是从提高原料产出效率的角度,肠衣加工企业亟盼依靠先进技术进行加工、工艺创新,以提高劳动生产率,增强企业的核心竞争力.

2 问题的提出

某在国内具有一定影响的肠衣加工企业希望能够借助计算机,计算速度快、结果可靠的特点,开发一套计算机软件。该软件一端线上人员测量所得原料数据输入电脑,就能自动生成经过优化后的满足成品规格要求的搭配方案。这样既可以减少劳动强度,又能提高取和使率。企业负责人有海外留学背景,认识到这个程序的核心是要建立原料搭配优化的数学模型,因此和本文作者进行了接触。

在首次接触中,我们的第一反映是这个问题最终需要归纳成一个优化问题<sup>[1]</sup>,具体是何种问题则需要作进一步的了解。为此我们需了解两种信息:原料信息和成品信息。

肠衣原料不同于一般的刚体,其具有一定的伸缩性,企业的测量以0.5米为一档,如:3.1米、3.5米按3米计算,3.6米、4米按3.5米计算,其余的依此类推(这样测量的约束是每根原料的测量长度均小于等于其实际长度,企业实际上是以部分原料的免费水运来保,成品的总长度的合格)。以下为某批次原料的测量结果,见表1。

表1: 原料描述表

长度(米)	3.1~3.5	3.6~4	4.1~4.5	4.6~5	5.1~5.5	5.6~6	6.1~6.5	6.6~7
根数	43	59	39	41	27	28	34	21
长度(米)	7.1~7.5	7.6~8	8.1~8.5	8.6~9	9.1~9.5	9.6~10	10.1~10.5	10.6~11
根数	24	24	20	25	21	23	21	18
长度(米)	11.1~11.5	11.6~12	12.1~12.5	12.6~13	13.1~13.5	13.6~14	14.1~14.5	14.6~15
根数	31	23	22	59	18	25	35	20
长度(米)	15.1~15.5	15.6~16	16.1~16.5	16.6~17	17.1~17.5	17.6~18	18.1~18.5	18.6~19
根数	30	42	28	42	45	49	50	64
长度(米)	19.1~19.5	19.6~20	20.1~20.5	20.6~21	21.1~21.5	21.6~22	22.1~22.5	22.6~23
根数	52	63	49	35	27	16	12	2
长度(米)	23.1~23.5	23.6~24	24.1~24.5	24.6~25	25.1~25.5	25.6~26		
根数	0	6	0	0	0	1		

对同一批原料,成品一般分成一种规格,每种规格用4个参数加以描述:最短原料长度,最长原料长度,原料根数,总长度。表2为对应于该批次原料的成品规格表。其中第一种大规格的最长原材料的长度记为 $\infty$ ,表示最长原材料的长度没有上限。

表2: 成品规格表

类别	最短长度(米)	最大长度(米)	根数	总长度(米)
小规格	3	6.5	20	89
中规格	7	13.5	8	89
大规格	14	$\infty$	5	89

基于上述信息,我们不难得到:

**目标1** 对于给定的一批原料,装+的衣+捆数越多越好

直觉告诉我+捆+包含1根+度为89米原料+成品和另+捆+89根长度为1米的原料组成的成品+比,前者的价值明显要高+因此+企业提+了+标2

**目标2** 对于成+捆数+的方案+最优+度最+的成品越多+方案越好

为了保证食品原料的新鲜,搭配方案的生成时必须足够快,经过操作部门的沟通,得到+标3.

**目标3** 为了食品保鲜,要在30分钟内+方案

从数学角度看+批原料使用+的本质是+问题+行域,也就是+成品规格表+档规格对于原料的限制+其中,原料根数和总+度+以做适当的放松,并从等式约束改为不等式,于是有:

**条件1** 总长度允许误差+围为+0.5-0.5+总根数允许误差+围为+1-0+

最+长度+关系到+成+质量+,能提高不能下降,但+提高会降低原料使用率,所以只好维持不变,作为+、+要+,-度+减+希望最大+度越大越好

**条件2** 某种规格对+原料+果+能解+,可以+级使用

如+变为14米的+和+以和+度大于7米的+进行捆扎,+成+属于7米-13.5米规格

### 3 问题的分析

首先看目标2+了+那里有两个难+第一+表述+似乎是一个最小+大+问题,但+实际+解+从+难+第二+,这个要求和+标1+似乎是+行+的,也就是满足+标1的方案可能不满足+标2,反之亦然

其次+,企业在+时+也+知道+了两个+标+矛盾+未解+的复杂性,所以已经“聪明+”地把成品规格+成+可+,+档+各档+成+生+论其最+下+度如何+律+等+其+实质+是将“最+下+度最+,+成+最多”的要求+化为“最+,存在某个值+的+成+最多”+这样+将一个从理论+为+优美但+无法+实现+的+标+转化为+近似“+解”+行为+行的+优化+和+

+标1+虽然简单,但+受目标2的+制+,+直接建+,+全部原料的+优化+模型,而+应该+合+条件+按照+种+的+规格+多+进行+优化+,时+为了提高原料使用率,可以+合+条件2,扩大每种规格最大长度的+限,

综合以上思考,我们给+求解+衣+装+配+问题的+个总体方案

- 1) 根据+规格要求,求最优解,
- 2) 将第1步优化后+余的原料纳入+规格+求最优解
- 3) 将第2步优化后+余的原料纳入+规格+求最优解
- 4) 如果+余的原料总+小于89.5米,则优化成功

以上总体方案的+特点是+用+个+优化+变+整体的+优化+每次优化的数学模型、求解方式相同+不同的是每+优化的+始+数据+原料+长度、根数和成品规格+此带来的好处是+,+需+编+个+函数,反复调+次,+间+高对原料+成+等+数据+作+必要的+调整

## 4 模型一

### 4.1 搭配方式模型

对于+种给+的规格,我们+min表示+原材料的最短长度,max表示+原材料的最+长度,number表示+最多使用的+材料根数+length表示+成品+长度+由于+原材料的长度是以0.5递

增的。那么,在  $\min$  和  $\max$  之间就总共有  $N = \max - \min + 0.5 + 1$  种不同长度的原材料,它们的长度记  $l_1, l_2, \dots, l_N$ 。材料  $i$  的根数记为  $y_1, y_2, \dots, y_N$ 。

搭配为  $x_1$  根表示为  $n_1, n_2, \dots, n_N$  表示,该种搭配方式中长度为  $l_i$  的原材料根数为  $n_i$ 。因为原材料根数为  $\text{number}$  或  $\text{number} - 1$ , 所以

$$\text{number} - 1 \leq \sum_{i=1}^N n_i \leq \text{number}$$

同时,由于成品的长度为  $\text{length} \pm 0.5$  或  $\text{length}$ , 即

$$\text{length} - 0.5 \leq \sum_{i=1}^N l_i n_i \leq \text{length} + 0.5.$$

结合实际原料的限制,就可以得出该规格所有搭配方式  $n_1, n_2, \dots, n_N$  所要满足的条件

$$\begin{cases} \text{number} - 1 \leq \sum_{i=1}^N n_i \leq \text{number} \\ \text{length} - 0.5 \leq \sum_{i=1}^N l_i n_i \leq \text{length} + 0.5, \\ 0 \leq n_i \leq y_i, \quad i = 1, 2, \dots, N \end{cases} \quad (1)$$

由于  $n_i$  是非负整数,所以满足上述条件的非负整数解的个数是有限的。而且,该问题快速求解满足上述条件的非负整数解,每个解对应一种搭配方式。

## 4.2 最优搭配模型

设  $i = 1, 2, \dots, M$  个解的个数为  $M$ 。其中第  $j$  个解为  $(a_{1j}, a_{2j}, \dots, a_{Nj})^T, j = 1, 2, \dots, M$ , 该解的搭配方案:第  $j$  种搭配方式,有  $x_j$  捆,那么,第  $j$  种搭配方式消耗的第  $i$  种原材料为  $a_{ij} x_j$ 。

2.  $N$ 、最优方案将消耗第  $i$  种原材料  $\sum_{j=1}^M a_{ij} x_j$  捆。而第  $i$  种原材料  $l_i$  有  $y_i$  根,所以消耗的根数不可能大于原材料数,即

$$\sum_{j=1}^M a_{ij} x_j \leq y_i \quad (2)$$

所以

$$\begin{cases} \sum_{j=1}^M a_{1j} x_j \leq y_1, \\ \sum_{j=1}^M a_{Nj} x_j \leq y_N. \end{cases} \quad (3)$$

所以

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1M} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \dots & a_{NM} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} \leq \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad (4)$$



记

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1M} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NM} \end{pmatrix},$$

则上式可写为

$$A\bar{x} \leq \bar{y}. \quad (5)$$

这样就确定了关于  $x_j$  的约束条件.

它的目标是能够得到尽可能多的成品或使剩余的原材料尽量少. 先考虑得到尽可能多的成品的要求. 为了得到尽可能多的成品, 须每种小规格的个数之和要最大, 也就是  $\sum_{j=1}^M x_j$  要最大. 这自然构成目标函数. 又因为实际要求,  $x_j$  为非负整数, 这样就得到一个整数规划如下:

$$\max \sum_{j=1}^M x_j \quad (6)$$

$$\begin{cases} A\bar{x} \leq \bar{y}, \\ \bar{x} \text{ 为非负整数向量} \end{cases}$$

再考虑使剩余的原材料尽量少. 记  $N$  阶列向量

$$\bar{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}$$

使得  $A\bar{x} + \bar{b} = \bar{y}$ , 则  $b_i$  就是第  $i$  种原材料的剩余长度. 要使剩余的原材料尽量少, 即要使  $\sum_{i=1}^N b_i$  最小. 这样得模型:

$$\min \sum_{i=1}^N b_i \quad (7)$$

$$\begin{cases} A\bar{x} + \bar{b} \leq \bar{y}, \\ \bar{x}, \bar{b} \text{ 为非负整数向量} \end{cases}$$

因为原材料的总长度是一定的, 设为  $C$ . 总长度分为两部分, 一部分用来做成品, 消耗  $89 \sum_{j=1}^M x_j$ , 另一部分被剩余, 剩余的总长度为  $\sum_{i=1}^N b_i$ , 则可得

$$89 \sum_{j=1}^M x_j + \sum_{i=1}^N b_i = C.$$

所以  $\sum_{j=1}^M x_j$  取到最大值时  $\sum_{i=1}^M b_i$  也取到最小值。因此, (6) 与 (7) 是等价的。

由于 (6) 只有  $M$  个决策变量, 而 (7) 的决策变量是  $M+N$  个, 从理论上讲, (6) 比 (7) 的求解速度要快, 因此, 通常采用前者来进行求解。但实际上, 由于  $M$  比  $N$  大很多, 因此,  $M$  和  $M+N$  相差也不是很大, 两者的计算速度也差异不大。

### 4.3 搭配方式模型求解

由于成品规格对  $w$  原料档次数不是常数, 搭配方式的生成无法使用循环方式来完成, 因此必须通过递归处理。而且原材料的长度是以 0.5 递增的, 为了加快运算速度, 可以把长度都乘以 2 转换成整数来处理。那么, 相邻两种长度的肠衣的长度之差为 1。

如果采用 Matlab 编程, 递归所需运行时间较长, 很可能导致总体计算时间超过 30 分钟的限制, 因此一般采用 C 程序生成小规格, 这样生成搭配方式可以在几分钟甚至 1 s 秒内完成。

### 4.4 优化模型求解

本文采用 Lingo 来求解整数规划<sup>[2,3]</sup>。搭配方式保存在 txt 文件中。虽然 Lingo 可以导入 txt 文件, 但直接导入时是逐行导入, 而对于  $M$  比较大的情况时, 分行导入显然是不合适的。为了接合 Lingo 中输入小规格情况数据也是不现实的。因为 Lingo 可以直接导入 Excel 文件, 因此本文先将 txt 文件在 Excel 中打开, 并加入原材料根数的数据, 再定义两个名称, 保存为 Excel 文件之后, 导入到 Lingo 中。同时, 在求解整数规划之前, 本文还将最优解输入到 txt 文件中, 以方便查阅。

第一步 下面是求解大规格搭配方案整数规划的 Lingo 代码:

```
model
sets
rows/1..2862/:x;
cols/1..23/:y;
link(rows,cols):A;
endsets
max=@sum(rows(i):x(i));
@for(cols(j):
@sum(rows(i):A(i,j)*x(i))=y(j));
@for(rows(i):@gt(x(i),
data
A=@ole('mydata3.xlsx','kinds');
y=@ole('mydata3.xlsx','y');
@text('mysol3.txt')=rows,'最优解' x;
enddata
end
```

下面给出 Lingo 代码。在  $M=2862$  的情况下, 最优目标值为 137, 最优解如表 3 所示。表上每行表示一种搭配方式, 行中最右一个数字表示该种搭配对  $w$  的捆数。

查表剩余, 只剩余一根长度为 18 的, 因此这样的组装方案显然是最优的。

第二步 将上面多余的原料降级到中规格进行搭配, 可算得  $M=19635$ , 最优捆数为 37 捆, 详细情况如表 4 所示。剩余原料如表 5 所示。

尽管剩余材料总长度超过 89.5, 但是用 5.1 所述捆数上限模型可以算得中规格捆数的上限为 37 捆, 和本算例所得相等。

表3 大规格搭配方案

		原料长度(米)																										根数
		14	14.5	15	15.5	16	16.5	17	17.5	18	18.5	19	19.5	20	20.5	21	21.5	22	22.5	23	23.5	24	24.5	25	25.5			
搭 配 方 案			1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	
			0	0	0		0	0	0		0		0	1	0	0	1	0	0	0	0	0	0	0	0	4		
			0	0	0	0	1		0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2		
			0	0	0	0	1	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	0	0	7		
			0	0	0	0	0	0	1	0	1	1	1	3	0	0	0	0	0	0	0	0	0	0	0	8		
	0			0	0	0	0	0	1		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	15		
	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10		
	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	8		
	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	16		
	0	0		0	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	6		
	0	0	0	1		1	0	0	0	0	0	0	1		0	1	0	0	0	0	0	0	0	0	0	10		
	0	0	0	1	0	0	1	0		1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	11		
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	1	0	0	0	0	6		
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	2		
	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	18		
	0	0	0	0	0	1		1	1	0		0	1	0	0	0	0	0	0	0	0	0	0	0	0	17		
0	0	0	1		0	0	1	0	0		0	0	0	0	1	0	0	0	0	0	0	0	0	0	1			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0		4			

表4 中规格搭配方案

		原料长度(米)															捆数
		7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14	
捆 配 方 案		0	0	1	1	0	0	0	0	1	0	0	0	3	0	0	4
		0	0	1	1	0	1	0	0	0	3	0	3	0	0	0	2
		0	0	1	0	3	0	0	0	0	0	0	0	1	3	0	1
		0	0	1	0	0	3	4	0	0	0	0	0	0	3	0	1
		0	0	1	0	0	0	0		2	0	4	0	0	0	0	4
		0	0	0	2	1	0	0	0	0	0	2	2	0	1	0	2
		0	0	0	1	1	0	0	2	0	2	0	0	0	2	0	5
		0	0	0	1	0	2	0	0	1	0	0	4	0	0	0	11
		0	0	0	0	4	0	0	0	0	3	0	0	0	0	1	1
		0	0	0	0	1	0	0	4	0	0	0	3	0	0	0	1
		0	0	0	0	0	0	4	0	1	1	0	0	2	0	0	4
		0	0	0	1	0	0	1	0	2	0	2	2	0	0	0	1

第三步 将上述多余原料纳入小规格原料进行优化搭配时,即使不考虑长度和根数的变化区间,只考虑严格相等约束条件,  $M$  也近 500 万个解,难以想象在微机上可以实施计算,必须将求解理论最优转化为求可行的最优(次最优解),对于企业也是可以接受的。得到次最优解的

主要方式是缩小搜索空间,如我们可以等步长地从32个搭配方式中选取Lingo所允许的最大个数作为搜索子空间进行优化。对本例数据,我们得到需用Lingo作优化的24564个搭配方式,最终得到目标函数值为16(如表6所示)。剩余原料情况如表7所示。

尽管剩余材料总长度超过80%,但是同5.1所述捆数上限模型可以算得小规格捆数的上限为16捆,所以所得近似最优可以认为是全局最优的。

表5: 中规格原料剩余情况

长度(米)	7	7.5	8	9.5	13	13.5
根数	24	24	8	1	1	

表6: 小规格近似优化搭配方案

	原料长度(米)														捆数
	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	9.5	13	13.5	
搭 配 方 式	1	0	1	1	1	4	1	0	1	3	0	0	1	0	1
	2	4	4		0	2	0	1	0	0	1	1	0	0	1
	3	4	3	0	4	0	0	1	1	1	0	0	0	3	1
	4	3	2	4	2	0	0	4	3	0	0	0	0	0	1
	5	1		2	1	4	0	3	1	1	0	0	0	2	
	6	3		0	1	3	2	0	1	0	1	0	0	1	1
	7	2	3	3	0	3	1	0	1	0	0	0	0	3	
	8	4	3	8	1	1	1	1	0	0	1		0	0	3
	9	0	0	0	8	0	1	6	1	0	0	0	0		
	10		1	1	0	0	1	2	0	8	0	0	0	0	1
	11	0	1	1		0	7	0	5	6	0	0	0	0	1

表7: 小规格原料剩余情况

长度(米)	3	3.5	5.5	7
根数	27	1	2	1

## 5 模型二

模型一先求所有符合条件的所有搭配方式,再根据这些方式展开全局优化,所得到的最优解也是实际需要的最优解,其不足在于:在进行小规格优化搭配时,因为搭配方式过多而只能得到局部最优解。

事实上,因为一批次原料根数不会很大,因此成品数量远远小于可供选择的搭配方式数量。由此我们得到解决问题的另外一种思路:模仿人工搭配方式,将最优捆数和搭配方式一起作为优化变量进行求解。具体做法也分两种:

1) 先估计出成品件数,再求出可能的搭配方式.

2) 直接求出最优捆数和搭配方式.

### 5.1 捆数上限模型

对符合某种规格 $n$ 的原料,设  $\text{number} = \text{length} \times N$ ,  $y_j$  ( $j = 1, 2, \dots, N$ ) 和模型 (5) 的意义相同,  $M$  为该规格成品数量,  $x_j$  为这些成品中第  $j$  种原料被使用的总数,  $j = 1, 2, \dots, N$ , 则  $M$  满足

$$\max M \quad (8)$$

$$\begin{cases} M \cdot \text{length} - 0.5 \leq \sum_{j=1}^N x_j l_j \leq M \cdot \text{length} + 0.5 \\ M(\text{number} - 1) \leq \sum_{j=1}^N x_j \leq M \cdot \text{number}, \\ x_j \leq y_j, \quad j = 1, 2, \dots, N \\ M, x_j \text{ 为非负整数} \end{cases}$$

模型 (8) 参数信息明确, 个数无约束, 可以直接调用 Lingo 软件求解得到最优解  $M^*$ . 进一步, 设  $x_{ij}$  为第  $i$  捆成品中第  $j$  种原料的捆数,  $i = 1, 2, \dots, M_0$ ,  $j = 1, 2, \dots, N$ , 则满足

$$\begin{cases} \text{length} - 0.5 \leq \sum_{j=1}^N x_{ij} l_j \leq \text{length} + 0.5, \quad i = 1, 2, \dots, M \\ \text{number} - 1 \leq \sum_{j=1}^N x_{ij} \leq \text{number}, \\ \sum_{i=1}^{M_0} x_{ij} \leq y_j, \quad j = 1, 2, \dots, N, \\ x_{ij} \text{ 为非负整数} \end{cases} \quad (9)$$

### 5.2 捆数和捆扎方案联合求解模型

设  $\text{number}$ ,  $\text{length}$ ,  $N$ ,  $l_j$ ,  $y_j$  ( $j = 1, 2, \dots, N$ ) 和模型 (5) 意义相同,  $M$  为成品数量,  $x_{ij}$  为第  $i$  捆成品中第  $j$  种原料的捆数,  $i = 1, 2, \dots, M$ ,  $j = 1, 2, \dots, N$ , 则  $M$  和  $x_{ij}$  满足

$$\max M \quad (10)$$

$$\begin{cases} \text{length} - 0.5 \leq \sum_{j=1}^N x_{ij} l_j \leq \text{length} + 0.5, \quad i = 1, 2, \dots, M \\ \text{number} - 1 \leq \sum_{j=1}^N x_{ij} \leq \text{number}, \\ \sum_{i=1}^M x_{ij} \leq y_j, \quad j = 1, 2, \dots, N \\ M, x_{ij} \text{ 为非负整数}, \end{cases}$$

## 6 模型分析

模型 (8) 的变量个数为  $M_0 \times N$ , 比模型 (6) 的变量个数  $M$  相比要小得多, 因此求解相对比较容易. 一般不会发生因为  $M$  太大而无法优化的情况. 但是, 在求解搭配方案问题 (9) 时, 因为  $M$  没有目标函数, 所以往往每次只能得到局部最优.

模型 10 可以直接通过 Lingo 求解, 但作为第一个约束条件的求和项数也是一个优化变量, 因此本质上它是一个非标准的整数规划问题, 即使使用 Lingo 也无法得到整体最优, 每次只能得到一个局部最优。

更值得指出的是, 在依据 (9) (10) 求解搭配方案的过程中, 一次优化只能得到一种搭配方式的最多组数, 为了得到全部材料的搭配, 需要修改剩余原料数据再次优化, 如此逐步进行, 直到剩余材料无法成捆为止。

总体而言, 模型 1 和 1 的搭配方式比较接近, 结果也容易理解, 但执行过程人工干预较多。一般不能保证在 30 分钟内得到搭配方案。模型 1 对问题的理解比较彻底, 是一个标准的整数规划模型, 理论上可以得到真正的最优解, 只要编程得当, 基本不需要人工干预, 符合企业的最终要求。不过由于该问题是 NP 困难的, 所以在搭配方式较多的情况下, 理论最优解无法在规定时间内求得, 必须通过广空间内搜索而得到一个局部最优解。

## 7 总结

作者之所以将此问题公开作为全国数学竞赛题, 是因为无论是通过模型 1 还是模型 1, 这个问题可以表达为一个数学规划问题, 而这对一个接受过数学建模训练的大学生而言应该不是一个难题。但是, 实际情况没有想象的那么简单。

首先, 相对 C 题, 选做本题的同学比较少。其中可能有两个原因, 第一, 学生已经习惯了通过网络搜集解题信息(这是当代大学生应该具备的技能), 但是“肠衣搭配优化问题”是一个源于中国的“土”问题, 网络上除了肠衣厂商的广告网站, 几乎得不到任何有用的资料, 许多参赛同学因此心有余悸, 不敢尝试。第二, 也有一些同学感觉它和经典的“钢材切割问题”非常相像, 只不过前者是化整为零, 后者是集零成整, 应该容易求解。但是, 这表面的细微差别却导致问题的本质发生了变化, 使变量个数成倍增加, 以致无法在规定时间内, 现有的软件完成计算, 他(她)们也为此而放弃了努力。对于第一类同学, 作者认为在数学建模竞赛过程中, 模型固然是一个核心, 但是有了模型如何求解, 也是解决实际问题的一个重要环节。在计算机软硬件条件十分成熟今天, 我们当然不必事必躬亲, 自己设计算法, 自己编程求解, 而应采用充分利用国内外先进的计算软件。但是, 实际问题中会出现计算规模超过, 算力能力的情况, 一种可行的解决办法就是将漂亮的理论最优解代之以实用的局部最优解。从此次阅卷的情况来看, D 题获奖的参赛队, 都或多或少地应用了这个思想。这也提醒了我们教师在教学过程中应该适当增加近似计算方面的内容。

其次, 本次做 D 题的绝大多数参赛队采用的建模思路是第一种模型, 没有用模型 1。其中原因我们在前面也分析过。它的思路和人 1 操作比较接近, 有的参赛队甚至在论文中指出将可+产线的工人请教过。相比之下, 模型 1 思路不是那么直接, 虽然它是经典的整数规划模型, 而求解的可靠性更高, 也更加容易改编成企业需要的软件产品。因此, 通过本题, 参赛同学会对整数规划模型及其计算有更深层的理解, 这对于将来解决实际问题是有帮助的。

最后, 要特别感谢谭永华、叶其孝、蔡志本等教授在 2010 年重庆工业工程数学研讨会上的启发和帮助, 也要感谢全国大学生数学建模竞赛组委会及专家组在出题过程中所给予的指导。

## 参考文献:

- 1 胡运权. 运筹学(第3版)[M]. 北京: 清华大学出版社, 2006.  
Hu Y Q. Operations Research (3rd Edition)[M]. Beijing: Tsinghua University Press, 2006.

- 2) 刘璟思 林 Lingo 在求解组合优化问题[J] 现代计算机(专业版) 2005, (6): 81-82  
Liu J Z. Solving optimization of combination problem based on Lingo[J] Modern Computer 2005, (6), 81-82
- 3) 丁卫东 姚志刚 程高 Lingo 语言与 0-1 集合整数规划模型的再结合[J] 物流工程与管理 2009, (3): 10-72-75  
Ding X D, Yao Z G, Cheng G. Lingo language in the logistics distribution center location[J] Logistics Engineering and Management, 2009, 31(10): 72-75

## Natural Casing Assembling Optimization: Models and Computation

LU Li-qiang<sup>1</sup>, YU Bing-quang<sup>1</sup>, FAN An<sup>2</sup>

(1 School of Mathematical Sciences, Fudan University, Shanghai 200433

2 China Unionpay Co., Ltd, Shanghai 200135)

**Abstract** This paper focuses the problem of optimization in natural casing assembly. Based on the thorough study on its background and requirements, we propose two mathematical models and analyse their computing behavior in integer programming.

**Keywords:** natural casing assembling; integer programming; mathematical modeling





## 2 模型假设

- 1) 假设原料最多只能降一级使用;
- 2) 假设每捆长度的原料都以0.5米为一档,如 3-3.5米,按3米计算;
- 3) 假设每次肠衣制作加工一分所发的原料对应长度与根数统计是正确的;
- 4) 假设本次的肠衣制作加工后剩余的原料不能用于下次的组装.

## 3 符号说明

- $m_i$  表示第  $i$  档长度原料的长度,  $i = 1, 2, \dots, 46$
- $x_{ij}$  表示第  $j$  捆成品中第  $i$  档长度原料的根数,  $i = 1, 2, \dots, 46, j = 1, 2, \dots, n$
- $p_i$  表示第  $i$  档长度原料的根数
- $L$  表示每捆成品的总长度
- $n_k$  捆去的方案
- $b_k$  表示3米到第  $k$  规格最长长度原料的入度档数
- $r_i$  表示原料不降级使用时第  $i$  档长度原料组装  $n_k$  捆后剩余的根数
- $q_k$  表示第  $k$  规格成品中每捆的根数,  $k = 1, 2, 3$
- $r_{2k}$  表示原料不降级使用时第  $k$  规格原料组装为成品  $Q_1$  捆数
- $Q_1$  表示规格2剩余的原料降级  $n_1$  与规格1剩余的原料组装为成品捆数
- $Q_2$  表示规格3剩余的原料降级  $n_2$  与规格2剩余的原料组装为成品捆数
- $y_{ij}$  表示原料降级使用时第  $j$  捆成品中第  $i$  档长度原料的根数
- $M$  表示最短长度最长的成品捆数

## 4 模型建立与求解

### 4.1 模型的建立

考虑上述建模的办法,首先建立3种规格产品组装的最多捆数为4标,即建立满足要求(1),(3),(4)的优化模型.其次,基于第一步模型求解结果,建立满足要求(2)的优化模型.

#### 4.1.1 基于要求(1),(3),(4)的模型建立

为了提高原料的使用率,每捆产品的总长度允许有0.5米的误差、总捆数允许比标准少1根并且原料可以降级使用(假定只能降一级,规格1对应的原料不能降级使用).从而以各种规格产品的最大捆数之和为目标函数建立整数规划模型.

目标函数为

$$\max Z = \sum_{k=1}^3 n_k + Q_1 + Q_2,$$

约束条件为

$$I - 0.5 < \sum_{i=y_{k-1}+1}^{y_k} m_i x_{ij} < I + 0.5, \quad k = 1, 2, 3, \quad j = 1, 2, \quad n_k \quad (1)$$

$$g_k - 1 \leq \sum_{i=y_{k-1}+1}^{y_k} x_{ij} \leq g_k, \quad k = 1, 2, 3, \quad j = 1, 2, \dots, n_k, \quad (2)$$

$$p_k - \sum_{j=1}^{n_k} x_{ij} = v_k, \quad k = 1, 2, 3, \quad j = 1, 2, \dots, n_k, \quad (3)$$

$$L - 0.5 < \sum_{i=y_{k-1}+1}^{y_k+1} m_i s_{ij} < L + 0.5, \quad k = 1, 2, 3, \quad j = 1, 2, \quad n_k, \quad (4)$$

$$g_k - 1 < \sum_{i=y_{k-1}+1}^{y_k+1} s_{ij} < g_k, \quad k = 1, 2, 3, \quad j = 1, 2, \quad n_k, \quad (5)$$

$$\sum_{j=1}^{Q_1} s_{ij} \leq v_i, \quad i = 1, 2, \dots, y_1, \quad (6)$$

$$\sum_{j=1}^{Q_1+Q_2} s_{ij} \leq v_i, \quad i = y_1 + 1, \dots, y_2, \quad (7)$$

$$\sum_{j=1}^{Q_2} s_{ij} \leq v_i, \quad i = y_2 + 1, \dots, y_3, \quad x_{ij}, s_{ij}, v_i \in \mathbb{Z}^+ \quad (8)$$

其中约束 (1), (2) 分别表示要求 (3) 的成品总长度约束和总根数约束。约束 (3) 为每种规格的原料总根数减小到原料组装成成品的根数等于剩余的根数。约束 (4), (5) 分别表示降级使用的原料部分满足要求 (3) 的成品总长度约束和总根数约束。约束 (6), (7), (8) 分别为每种规格产品中, 降级使用的原料数量不能超过每种规格产品中降级前组装成成品中剩余的原料数量。

#### 4.1.2 增加考虑要求 (2) 的“优中选优”的模型建立

对于成品捆数相同的方案, 最长长度最长的成品越多, 方案越好, 即

$$\max \text{card} \{ \min \{ s_u \} \mid u = 1, 2, \dots, n, \text{ 且 } \min \{ s_u \} = M \},$$

其中  $\max \{ \min \{ s_u \} \} = M$ ;  $\text{card} \{ A \}$  表示集合  $A$  的元素个数;  $s_u$  表示捆法的方案,  $M$  表示最短长度最长的成品捆数。

#### 4.2 基于满足要求 (5) 的启发式算法求解

根据该模型的 NP 属性, 直接求解, 求解时间远远超过了题目要求 (5) 的规定时间, 所以需要构造启发式算法进行求解。

##### 4.2.1 启发式算法的构建

算法符号说明  $q_j$  表示第  $j$  步最多捆数的取法。

步骤1 初始化  $j = 1, k = 1$ , 求解下列模型:

$$\begin{aligned} \max \quad & q_j \\ \text{s.t.} \quad & \begin{cases} g_k - 1 \leq \sum_{j=p_{k-1}+1}^k q_j \leq g_k, \\ L - 0.5 \leq \sum_{j=p_{k-1}+1}^k m_j q_j \leq L + 0.5, \\ v_i = p_i, \quad q_j x_{ij} \geq 0, \\ q_j, p_i \in Z^+, \end{cases} \end{aligned}$$

步骤2 将结果进行选择, 即对最短长度最长的成品越多方案越好:

$$\max \{ \min \{ s_u \} \} = M, \quad \max \text{card} \{ \min \{ s_u \} \mid u = 1, 2, \dots, n, \text{ 且 } \min \{ s_u \} = M \}$$

得到一种最好的方案.

步骤3 将  $i$  换为  $p_i$  返回步骤1, 令  $j = j + 1$ . 当  $m_j$  无可行解或  $m_j = 0$  时, 终止  $j$ , 执行  $n_k = \sum_{j=1}^k m_j$ , 令  $k = k + 1$ .

步骤4 将步骤3上剩余的重新记为  $p_i$  再将  $i$  的求和范围变为  $p_{k-1} + 1$  到  $p_k$ , 转步骤1.

#### 4.2.2 启发式算法的求解结果

根据启发式算法, 求得了每种规格下的最大捆数情况下的最好方案. 规格1的方案见表1, 规格2的方案见表2, 规格3的方案见表3. 规格2降级使用的方案见表4; 规格3降级使用的方案见表5.

通过启发式算法, 综合考虑要求(1)-(4), 求得规格1的捆数为13捆, 规格2的为34捆, 规格3的为134捆, 规格2降级使用最多总捆数为6捆, 规格3降级使用最多总捆数为17捆. 求得总捆数为208捆. 我们还上该启发式算法用时为1318秒(小于30分钟), 所以该算法满足要求(5). 并且我们求得的启发式算法的捆法方案比较少, 即可操作性强.

表1: 规格1的方案

捆法	捆数	3	3.5	4	4.5	5	5.5	6	6.5
1	11	0	5	3	3	2	2	3	1
2	2	10	0	0	0	2	3	0	5
剩余		23	4	6	8	1	0	1	0

表2: 规格2的方案

捆法	捆数	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5
1	22	0	1	0	1	0	0	0	1	1	1	1	2	0	1
2	9	0	0	0	0	2	0	0	2	1	0	0	1	2	0
3	3	0	0	0	0	0	0	5	0	0	0	0	2	0	1
剩余		24	2	20	3	3	23	6	0	0	1	0	0	0	0

表3 規格3的才量

抽去	抽數	14	14 5	15	15 5	16	16 5	17	17 5	18	18 5	19	19 5	20	20 5	21	21 5	22	22 5	23	23 5	24	24 5	25	25 5
1	49	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	32	0	0	0	0	0	1	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	16	0	0	0	2	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
4	12	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
5	6	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0
6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
7	2	0	0	0	1	0	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
8	1	0	3	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
剩餘		35	26	3	2	0	0	1	0	1	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0

表 1 煙煤 2 降級使用

捆法	捆数	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5
1	5	4	0	1	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0
2	1	3	3	1	0	1	0	1	0	1	0	0	1	0	3	1	0	0	1	0	0	0	0
剩餘		0	1	0	3	0	0	0	0	0	2	0	2	3	0	0	0	0	0	0	0	0	0

表5 规格3降阻使用

[illegible]

### 4.3 改进的邻域整点搜索的启发式算法<sup>1</sup>

#### 4.3.1 有关概念

松弛最优解 指略去变量为整数的条件原 IP 问题(记为 B) 松弛问题、记为  $A_1$  的最优解, 即

$$\begin{cases} x^* = (x_1^*, x_2^*, \dots, x_n^*)^T, \\ Z^* = \sum_{j=1}^n c_j x_j^*. \end{cases}$$

松弛最优解邻域 简称邻域) 指满足如下条件的解  $x \in X$  的集合  $x_j^* < x_j \leq [x_j^*] + 1$

限界约束 分为上界约束  $x_j < [x_j^*] + 1$  和下界约束  $x_j^* \leq x_j$

弱化约束 指满足条件  $x_j^* < x_j$  的约束条件  $x_j \leq x_j^*$  (对求极大解), 或满足  $x_j^* > x_j$  的约束条件  $x_j \geq x_j^*$  (对求极小解).

#### 4.3.2 改进的邻域整点搜索的启发式算法

$$\left\{ \begin{array}{l} L - 0.5 \leq \sum_{i=y_k-1}^{y_k} m_{ij} x_{ij} < L + 0.5 \\ g_k - 1 < \sum_{i=y_k-1}^{y_k} x_{ij} < g_k \\ p_i - \sum_{j=1}^{n_k} x_{ij} = v_i \\ L - 0.5 \leq \sum_{i=y_k-1}^{y_k} m_{ij} s_{ij} < L + 0.5 \\ g_k - 1 < \sum_{i=y_k-1}^{y_k} s_{ij} < g_k \\ \sum_{j=1}^Q s_{ij} \leq v_i, \quad i = 1, 2, \dots, y_1 \\ \sum_{j=1}^{Q_1+Q_2} s_{ij} \leq v_i, \quad i = y_1 + 1, \dots, y_2 \\ \sum_{j=1}^{Q_2} s_{ij} \leq v_i, \quad i = y_2 + 1, \dots, y_3 \end{array} \right. \quad (9)$$

$$x_{ij}, s_{ij}, v_i \in Z^+, \quad (10)$$

$$\max \text{card}\{\min\{s_u\}, u = 1, 2, \dots, n \mid \min\{s_u\} = M\} \quad (11)$$

步骤 1 略去整数约束 (10) 求其松弛最优解, 即先给定  $n$  值,  $n$  从拥数上限取起, 如第  $n$  等级从上限为 14 拥取起, 如果无可行解,  $n$  减少 1 再代入 (7) 式找可行解, 直到找到可行解,

步骤 2 求得松弛解记为  $x_{ij}^*, s_{ij}^*$ ;

步骤 3 确定邻域的上下界  $[x_{ij}^* - x_{ij}^*] + 1, [s_{ij}^*] - [s_{ij}^*] + 1$ ;

步骤 4 列出邻域边界的整点集合:

$$x_1^* = (x_{1j}^*, s_{1j}^*, \dots, x_{2j}^* - x_{1j}^* - s_{1j}^* + 1), \quad x_3^* = ([x_{1j}^*] + 1, s_{1j}^*), \quad x_4^* = ([x_{1j}^*] + 1, s_{1j}^* + 1),$$

步骤 5 可行性检验, 将整点代入 (6) 式, 找出满足的集合, 通过检验就是原问题的近似最优解

步骤 6 代入 (8) 式取其中最好的解

### 4.3.3 基于改进的邻域整点搜索的启发式算法求解结果(略)

### 4.4 两种启发式算法的求解结果比较

第二种启发式算法的方案捆数见表6,改进的邻域整点搜索启发式算法的方案捆数见表7

由第一种启发式算法求得的最多总捆数为208捆,用改进的邻域整点搜索的启发式算法求得的总捆数为166捆,从结果可看出第二种启发式算法求得的结果更接近于最优解,其算法程序运行的时间为1117秒,改进的邻域整点搜索的启发式算法程序运行的时间为600秒,对两种算法的综合评价,邻域整点搜索算法的优点为操作简便,能较快的求解结果和方案,缺点是结果只是局部极优,第二种启发式算法的优点是解的值接近于捆数的上限,缺点是操作较繁琐

表6 对于5种规格的总捆数

规格1的 总捆数	规格2的 总捆数	规格3的 总捆数	规格2降级 使用总捆数	规格3降级 使用总捆数	3种规格 总捆数
13	34	134	6	17	208

表7 改进的邻域整点搜索的启发式算法的方案捆数

规格1的 总捆数	规格2的 总捆数	规格3的 总捆数	规格2降级 使用总捆数	规格3降级 使用总捆数	3种规格 总捆数
9	25	103	24	5	166

## 5 结果分析

此类问题是属 NP 难题,不存在多项式精确的算法,只能寻求一种近似算法,通过表2可知规格1的根数为292根,最多能捆14捆,余12根,总长度为 $1355 \times 10^3$ 米,最多能组成14捆,余59.5米,则规格1的上限为14捆,同理,规格2的上限为34捆,规格3上限为135捆,通过第一种启发式算法求得规格1的捆数为13捆,规格2为34捆,规格3为134捆,通过比较可知求得的结果与上限非常接近,所以该启发式算法是有效的,而且我们算法的运行时间为1118秒(不到30分钟),算法满足题意,并且我们求得的启发式算法的捆数方案比较少,可操作性强

### 参考文献:

- 郭永发 等. 邻域整点搜索法求解整数规划[J]. 数学的实践与认识, 2006, 36(11): 100-104  
Guo Y F, et al. Neighborhood integer point search method to solve integer programming[J]. Mathematics in Practice and Theory, 2006, 36(11): 100-104

## A Heuristic Method for Solving the Assignment Problem of Natural Casing Materials

YANG Zhong, ZHANG Qi-liang, XU Xiao-hui

Advisor: LING Wei-wei, LI Fan-chun

(Department of Basic Courses Jiangxi College of Applied Technology, Ganzhou 341000)

**Abstract.** In this paper, we establish an optimization model for the assignment problem of the natural casing, and design a heuristic algorithm to solve the problem. For the requirement 1, we reduce the problem into a maximal number of bundles for size 1, size 2 and size 3. Based on the Planning theory, we establish an optimization model for the problem in condition that the number and the length of each product are fixed and known, and also we get the maximal number of bundle is 182. For the requirement 2, we obtain the material number of each bundle of the scheme, which gets the maximal number of bundles by employing Lingo 10.0. As compared with all schemes, we obtain the best scheme which gets the maximal number of maximal length of shortest materials in each bundle. For the requirements (3 and 4), first of all, we obtain the number of each bundle size on the assumption of no degraded using, and then we get the bundle number of remainder materials on the assumption of degraded using. Finally we get the maximal bundle number 208 and obtain the optimal scheme. For the time limit requirement (5), based on the establishment of large-scale integer programming model, this paper introduces a creative heuristic algorithm and obtains the approximate solution. The programs of the proposed heuristic algorithm can attain results within 1118 seconds. Worker can arrange the production according to side III a prescription. Also, we give an improved neighborhood point search algorithm to solve this question. We further compare the results and complexities of the two proposed methods.

**Keywords:** casing assignment, optimization models, heuristic algorithm





#### 4.1 构成一捆成品的方案

要想求出符合给定条件的最大捆数, 首先必须要分析构成一捆成品的方案有多少种. 比如通过分析对于第I类产品可以从3米、3.5米、4米、...、到6.5米中任取几个构成总数为19或20根, 总长度为88.5米、89米或89.5米的方案个数. 据此, 为了得出构成一捆第I类成品的方案, 我们建立分类模型(一)如下:

分类模型(一)

$$\begin{cases} 88.5 \leq 3x_1 + 3.5x_2 + \cdots + 6.5x_8 < 88.9, \\ 19 \leq \sum_{i=1}^8 x_i \leq 20, \\ x_i \in N, \quad i = 1, 2, \dots, 8. \end{cases}$$

通过 Matlab<sup>1</sup> 编程得出该模型的解  $A$ , 有 90849 个, 即构成一捆第I类成品的方案有 90849 个. 类似分类模型(一) 为了得到构成一捆第II类、第III类成品的方案, 我们建立分类模型(二)和分类模型(三)如下:

分类模型(二)

$$\begin{cases} 88.5 \leq 7x_1 + 7.5x_2 + \cdots + 13.5x_6 \leq 89.5, \\ 7 < \sum_{i=1}^{14} x_i < 8, \\ x_i \in N, \quad i = 1, 2, \dots, 14. \end{cases}$$

分类模型(三)

$$\begin{cases} 88.5 \leq 14x_1 + 14.5x_2 + \cdots + 22.5x_4 \leq 89.5, \\ 4 \leq \sum_{i=1}^{24} x_i \leq 5, \\ x_i \in N, \quad i = 1, 2, \dots, 24. \end{cases}$$

通过 Matlab<sup>1</sup> 编程程序类似分类模型(一), 得出分类模型(二)的解  $B$ , 有 10866 个, 分类模型(三)的解  $C$ , 有 4286 个.

#### 4.2 每类成品的最大捆数模型

从 4.1 的结果可知, 对于每类产品, 构成一捆的方法最少的都达到 4286, 若对所有的构成方案均进行考虑, 则运算量就会太大, 不易在 30 分钟内得出最多捆数的结果, 因此我们对这些构成方案进行筛选. 将构成此方案所需的原料种类超过一个的均舍去, 例如在构成一捆第I类成品的方案  $3 \times 1 + 3.5 \times 4 + 4 \times 3 + 4.5 \times 4 + 5 \times 6 + 5.5 \times 1 + 6 \times 1 = 88.5$ , 此方案包含了 3、3.4、3.5、3.9、4、4.4、4.5、4.9、5、5.1、5.5、5.9、6、6.4 共七种原料, 故将此方案舍去. 此假设在实际加工中也是合理的, 因为构成一捆肠衣的方案所需的长度范围越多其工作量就越大. 利用此假设可降低实际的工作量.

基于上述假设, 我们对 4.1 中得出的结果进行筛选, 得出构成第I类、第II类、第III类成品的方案分别有 668、439、792 种, 具体结果见表 1、表 2 和表 3.

我们先忽略第四个条件, 即在原料不可降级使用条件下的最大捆数. 由第一个假设可知要使总捆数最大, 实际上是三类成品在现有原料的条件下各自最大捆数的和, 故我们先分别对三类模型求出最大捆数. 下面我们以前第I类成品为例建立模型, 第II类、第III类产品则类似.

筛选后能够构成一捆第I类产品的方案有 668 种. 而该成品的原料限制, 如表 1.

现假设第  $i$  种构成一捆的方案  $A_i = (a_{i1}, a_{i2}, \dots, a_{i8}) (i = 1, 2, \dots, 668)$  有  $t_i$  捆, 各长度范围根数为  $o = (o_1, o_2, \dots, o_8) = (43, 59, 39, 41, 27, 28, 34, 21)$ , 则可知捆数  $t_i$  需满足各长度范围的

根数总数应少于  $a_j$ , 如长度范围位于 3-3.4 的应满足  $\sum_{i=1}^{668} a_{ji}x_i < a_j$ 。由此可知, 要在满足原料的条件下捆数最大, 实际是一个求最大的整数规划问题<sup>[2-4]</sup>。现建立第 I 类成品的最大捆数模型如下:

$$\begin{aligned} \max \quad & \sum_{i=1}^{668} L_i \\ \text{s.t.} \quad & \sum_{i=1}^{668} L_i a_{ij} \leq a_j, \quad j=1, 2, \dots, 8, \\ & L_i \in N, \quad i=1, 2, \dots, 668 \end{aligned}$$

该模型可以通过用 Lingo 编程<sup>[5]</sup>求解, 该程序的运行结果见表 2。表 2 中非 0 变量  $x_i$  所对应的方案  $A_i$  的具体数值, 见表 3。

表 1 第 I 类产品的原料

长度	1-3.4	3.5-3.9	4-4.4	4.5-4.9	5-5.4	5.5-5.9	6-6.4	30.5-6.9
根数	43	59	39	41	27	28	34	21

表 2 第 I 类成品的最大捆数模型的解

变量	$x_7$	$x_{131}$	$x_{227}$	$x_{294}$	$x_{431}$	$x_{462}$	$x_{473}$	$x_{514}$	$x_{532}$	$x_{615}$	其他
结果	2	1	1	1	1	2	1	1	1	3	0

表 3 第 I 类成品的最大捆数模型中非 0 解所对应的形成一捆产品的具体方案

方案	1-3.4	3.5-3.9	4-4.4	4.5-4.9	5-5.4	5.5-5.9	6-6.4	30.5-6.9
$A_7$	0	0	4	15	1	0	0	0
$A_{131}$	3	0	0	1	15	0	0	0
$A_{227}$	0	4	8	0	0	7	0	0
$A_{294}$	7	2	0	0	0	11	0	0
$A_{431}$	6	0	0	0	7	0	6	0
$A_{462}$	9	0	0	3	0	0	8	0
$A_{473}$	1	9	0	0	0	0	9	0
$A_{514}$	0	0	16	0	0	0	3	1
$A_{532}$	7	0	0	0	0	10	0	2
$A_{615}$	0	13	1	0	0	0	0	6

由表2和表3的结果可知第I类成品的最大捆数为14捆,此时原料的应用情况见表4.

表4 第I类成品的最大捆数模型中原料的应用情况与给出的原料数量的比较

长度	3-3.4	3.5-7.9	4-4.4	4.5-4.9	5-5.4	5.5-5.9	6-6.4	36.5-6.9
给出原料根数	43	59	39	41	27	28	34	21
使用根数	42	54	36	37	24	28	34	21

由表4可知,第I类成品的最大捆数模型中对于给出的原料最终只剩下16根,不能构成捆,因此第I类成品的最大捆数为14捆是合理的.类似第I类成品的最大捆数模型,对于第II类、第III类产品 $\mu$ 建立如下模型.第II类成品的最大捆数模型

$$\begin{aligned} \max \quad & \sum_{i=1}^{439} l_i \\ \text{s.t.} \quad & \sum_{i=1}^{439} l_i b_{ij} \leq p_j, \quad j=1,2,\dots,14, \\ & l_i \in N, \quad i=1,2,\dots,439 \end{aligned}$$

第III类成品的最大捆数模型

$$\begin{aligned} \max \quad & \sum_{i=1}^{792} l_i \\ \text{s.t.} \quad & \sum_{i=1}^{792} l_i c_{ij} \leq q_j, \quad j=1,2,\dots,24 \\ & l_i \in N, \quad i=1,2,\dots,792. \end{aligned}$$

其中 $p=(p_1,p_2,\dots,p_{14})=(24,24,20,25,21,23,21,18,31,23,22,59,18,25)$ 为第II类成品的原料数, $q=(q_1,q_2,\dots,q_{24})=(35,29,30,42,28,42,45,49,50,64,52,63,49,35,27,16,12,2,0,6,0,0,0,1)$ 为第III类成品的原料数.

对于第II类、第III类成品的最大捆数模型的求解算法类似于第I类成品的最大捆数模型,得到第II类成品最大捆数为37,对于给出的原料剩余58根.由于该模型求出的所有结果构成一捆第II类成品的方案中均未用到长度为7.5-7.9的原料,故该长度范围的原料全部剩余,而长度为7.7-4的也只用了一根,因此造成了第II类成品的原料剩余量较多.虽然如此,但长度位于8.5-13.4的原料完全用完,既长度较长的原料全部用完,这样有利于减少浪费,故第II类成品最大捆数模型也具有一定的合理性.

第III类成品得到的最大捆数为137,恰好是给出原料的根数.所有的原料被完全用完,很显然第II类成品最大捆数模型是最优的,各方案安排也是最合理的.

综合第I类、第II类、第III类成品的最大捆数模型,可知在不考虑降级情况下总捆数最大为188捆.虽然此时第III类成品的原料全部用完,但第I类、第II类的原料还有部分剩余,故模型有待改进.

## 5 模型的改进—综合最大捆数模型

第4部分给出了不考虑降级情况下的第I类、第II类、第III类成品最大捆数模型,得出最大捆数为188捆。但在这些模型中有很多材料被浪费掉,特别是长度为7.5-7.9的肠衣完全没有用到,7-7.4的肠衣虽然用到但也只用了一根,这对于大批生产的工厂来说损耗是很大的。为了减少材料的浪费,我们考虑将上一规格产品的材料剩余量降级使用。由第二个假设可知,原料降级使用时降到距离所在规格最近的长度范围,即第III类成品的剩余原料可以与第II类成品的原料进行捆扎,降级为13.5-13.9的原料使用。第II类成品的原料与第I类成品的原料进行捆扎,降级为6.5-6.9的原料使用。基于以上分析,我们可以求解第III类成品最大捆数模型,得出结果后看是否有剩余的原料,若无则正好分配完,继续解第II类成品最大捆数模型,若有剩余,则将剩余的原料作为13.5-13.9的原料,再求解第II类成品最大捆数模型,以此类推,直至求完第I类成品最大捆数模型,最后将三个模型的结果求和就得到最大捆数。此方法虽容易理解,但过程繁琐,因此我们需建立一个统一的模型来求出最大捆数。用 $x_i$  ( $i = 1, 2, \dots, 668$ )表示选择第 $i$ 种构成一捆第I类成品的方案 $A_i = (a_{i1}, a_{i2}, \dots, a_{i8})$ 的捆数, $y_i$  ( $i = 1, 2, \dots, 439$ )表示选择第 $i$ 种构成一捆第II类成品的方案 $B_i = (b_{i1}, b_{i2}, \dots, b_{i14})$ 的捆数, $z_i$  ( $i = 1, 2, \dots, 792$ )表示选择第 $i$ 种构成一捆第III类成品的方案 $C_i = (c_{i1}, c_{i2}, \dots, c_{i24})$ 的捆数。第I类产品有8种不同长度的原料3-3.4表示第1种原料,3.5-3.9为第2种原料,以此类推至第8种原料。类似地,第II类、第III类产品的原料分别为14、24种。第I类产品的第 $i$ 种原料在最优分配后的剩余量用 $d1_i$  ( $i = 1, 2, \dots, 8$ )表示,第II类产品的第 $i$ 种原料在最优分配后的剩余量用 $d2_i$  ( $i = 1, 2, \dots, 14$ )表示,第III类的第 $i$ 种原料在最优分配后的剩余量用 $d3_i$  ( $i = 1, 2, \dots, 24$ )表示,则考虑原料可以降级使用下的综合最大捆数模型为:

$$\begin{aligned} \text{MAX} \quad & \sum_{i=1}^{668} x_i + \sum_{i=1}^{439} y_i + \sum_{i=1}^{792} z_i \\ \text{s.t.} \quad & \sum_{i=1}^{668} a_{ij} x_i + d1_j = o_j, \quad j = 1, 2, \dots, 7, \\ & \sum_{i=1}^{668} a_{i,8} x_i + d1_8 - \sum_{i=1}^1 4d2_i = o_8, \\ & \sum_{i=1}^{439} b_{ij} y_i + d2_j = p_j, \quad j = 1, 2, \dots, 13, \\ & \sum_{i=1}^{439} b_{i,14} y_i + d2_{14} - \sum_{i=1}^{24} d3_i = p_{14}, \\ & \sum_{i=1}^{792} c_{ij} z_i + d3_j = q_j, \quad j = 1, 2, \dots, 24, \end{aligned}$$

$$x_i \in N, \quad i = 1, 2, \dots, 668, \quad y_i \in N, \quad i = 1, 2, \dots, 439,$$

$$z_i \in N, \quad i = 1, 2, \dots, 792, \quad d1_i \in N, \quad i = 1, 2, \dots, 8,$$

$$d2_i \in N, \quad i = 1, 2, \dots, 14, \quad d3_i \in N, \quad i = 1, 2, \dots, 24.$$



表6: 综合最大捆数模型中原料的应用情况与给出的原料数量的比较

长度	3-3.4	3.5-3.9	4-4.4	4.5-4.9	5-5.4	5.5-5.9	6-6.4	6.5-6.9
给出根数	43	59	39	41	27	28	34	21
使用根数	42	59	39	40	23	27	33	21
长度	7-7.4	7.5-7.9	8-8.4	8.5-8.9	9-9.4	9.5-9.9	10-10.4	10.5-10.9
给出根数	24	24	20	25	21	23	21	18
使用根数	15	24	10	25	21	23	21	18
长度	11-11.4	11.5-11.9	12-12.4	12.5-12.9	13-13.4	13.5-13.9	14-14.4	14.5-14.9
给出根数	31	23	22	59	18	25	35	29
使用根数	31	23	22	5	18	25	35	29
长度	15-15.4	15.5-15.9	16-16.4	16.5-16.9	17-17.4	17.5-17.9	18-18.4	18.5-18.9
给出根数	30	42	28	42	45	49	50	64
使用根数	30	42	28	42	45	49	50	64
长度	19-19.4	19.5-19.9	20-20.4	20.5-20.9	21-21.4	21.5-21.9	22-22.4	22.5-22.9
给出根数	52	63	49	35	27	16	12	2
使用根数	52	63	49	35	27	16	12	2
长度	23-23.4	23.5-23.9	24-24.4	24.5-24.9	25-25.4	25.5-25.9		
给出根数	0	6	0	0	0	1		
使用根数	0	6	0	0	0	1		

由表5可知,综合最大捆数模型所得的最大捆数为

$$\sum_{i=1}^{668} x_i + \sum_{i=1}^{439} y_i + \sum_{i=1}^{792} z_i = 191.$$

比分类最大捆数模型的总和要多3捆。由于篇幅有限,对于在情况最优下,各个 $x_i, y_i, z_i$ 所对应的具体构成一捆产品的方案在此不再列出,在表1、表2与表3中可查得相应的值。在综合最大捆数模型下得到第I类、第II类、第III类成品的捆数分别为17、37、137捆。我们会发现第I类成品的捆数比我们给出的材料所能构成的捆数要多,这就是原料降级使用的结果,因为我们的第II类成品原料有剩余,将其原料降为长度为6.5-6.9的原料使用,最终使得我们的第I类成品比分类最大捆数模型中多出3捆。实际上在表6中长度为7-7.4与7.5-7.9的原料的使用根数均是降为6.5-6.9所使用的根数,在本来的第II类成品最大中均未使用。

## 6 模型的总结与评价

本文在合理假设的基础上,首先通过建模得到构成一捆产品的全部方案。由于所得方案的总量过大,对于30分钟内得出分配方案无法做到,因此我们对方案进行筛选,舍弃跨越三个原料种类以上的方案,得出构成一捆第I类、第II类、第III类成品的方案分别有668、439、792种。对于筛选后得到的方案相比最初的数据减少了很多,而且在目前高科技飞速发展的时代,对于这个数据在30分钟内得出方案完全有可能,当然也可以在此基础上对方案再进行限制,得到更少的方案,但是这样有可能使某些长度范围的原料在任何一种方案中

都未使用到,那么此时对于那些未使用到的原料就非常的浪费,而在我们筛选后得出的方案中未出现这种状况,故筛选是合理的.在得出构成一捆成品的方案后,我们先建立在原料不可降级使用时的捆数最大模型.由原料不可分割的假设可知要想总捆数最大,实际上是三类成品在现有原料的条件下各自最大捆数的和,所以我们建立了分别对应三类产品的最大捆数模型,实际上也是含不等式约束的线性规划问题.第I类、第II类、第III类产品最大捆数模型下,我们得到它们的最大捆数分别为14捆、37捆、137捆.虽然第III类产品的原料全部用完,但第I类、第II类产品的原料还存在部分原料的剩余,三类产品分成三个模型,虽然能控制在30分钟内,但非常繁琐,而且没有考虑原料降级使用的条件,对原料的浪费较多,存在一些不合理的地方.由于分类最大捆数模型存在一系列的不足,所以我们在考虑原料可降级使用的条件下,得出一个全部产品一起考虑的模型,即综合最大捆数模型.该模型在之前分类最大捆数模型的基础上,加入一些松弛变量或人工变量,并对三类产品最大捆数模型进行综合,得出一个综合整数规划模型,求解得到最大捆数为191捆,比分类时多3捆.该模型达到最优解时,第III类产品的原料全部用完,故满足了最短长度最长的产品最多的条件.综合最大捆数模型既考虑了原料的降级使用,也能通过一个模型将总的最大捆数得出.虽然该模型的变量较多,但通过Lingo来求解整数规划模型速度很快,通过我们给出的程序,几分钟就得出了最终结果.因此对于一般情况的原料,在已得出的构成一捆产品方案的基础上,再应用我们给出的Lingo程序,在30分钟之内一定可以得出最大捆数方案.虽然最终的模型考虑到了所有的因素,但对于降级的问题我们只是将其降为下一等级长度范围最长的情况,对于是否可以降为其它长度未考虑,因此有待进一步的改进.

#### 参考文献:

- [1] 胡良剑,孙晓君. Matlab数学实验[M]. 北京:高等教育出版社, 2006  
Hu L J, Sun X J. Matlab Mathematical Experiment[M]. Beijing: Higher Education Press, 2006
- [2] 甘应爱,等. 运筹学(第3版)[M]. 北京:清华大学出版社, 2005  
Gan Y A, et al. Operations Research (3rd Edition)[M]. Beijing: Tsinghua University Press, 2005
- [3] 彭放,杨瑞康,等. 数学建模方法[M]. 北京:科学出版社, 2007  
Peng F, Yang R Y, et al. Mathematical Modeling Method[M]. Beijing: Science Press, 2007
- [4] 韩中庚. 数学建模方法及其应用[M]. 北京:高等教育出版社, 2005  
Han Z G. Mathematical Modeling Method and its Application[M]. Beijing: Higher Education Press, 2005
- [5] 谢金星,薛毅. 优化建模与Lindo/Lingo软件[M]. 北京:清华大学出版社, 2005  
Xie J X, Xue Y. Optimization Modeling and Lindo/Lingo Software[M]. Beijing: Tsinghua University Press, 2005

## Study on the Maximum Number of Bundles about Collocation of Natural Casing

ZHOU Ya, LIU Zhen-tang, LU Xue-yang

Advisor: HUANG De

(Guangxi City College, Guangxi 532100)

**Abstract:** This paper attains all product schemes of constituting a bundle of natural casing from the company's specific requirements. Because the scale of the attained schemes is too big, we abandon those raw materials more than three kinds of above schemes, and build the bundle number maximum model in the presumption that the raw material cannot be relegated to use. In order to improve the utilization rate of the raw materials, this paper considers that the raw materials can be degraded to use, and draw a general integer programming model for all the products. The optimal solution is calculated from all of the model solution bundles. The optimal solution is 191 bundles, and thus the optimum matching scheme can be determined.

**Keywords:** integer programming; classification scheme; Matlab; Lingo